DOCTORAL THESIS

2015

A COMPLEX NETWORK THEORY APPROACH TO OCEANIC AND ATMOSPHERIC TRANSPORT PHENOMENA

Enrico Ser-Giacomi

DOCTORAL THESIS

2015

Doctoral Program of Physics

A COMPLEX NETWORK THEORY APPROACH TO OCEANIC AND ATMOSPHERIC TRANSPORT PHENOMENA

Enrico Ser-Giacomi

Thesis Supervisor: Emilio Hernández-García
Thesis Co-Supervisor: Cristóbal Lopez

Doctor by the Universitat de les Illes Balears

Abstract

The last two decades have seen important advances in the Lagrangian description of transport and mixing in fluid flows driven by concepts from dynamical systems theory, and nowadays several approaches have been developed. Some of such techniques focus on geometric objects - lines, surfaces - separating fluid regions with different properties while others have focussed on computing stretching-like fields in the fluid domain, such as different types of Lyapunov exponents or other Lagrangian descriptors. Finally, there is a line of research focussing on the moving fluid regions themselves, the so-called set-oriented methods.

On the other hand many real-world systems can be studied by using the Network paradigm and in the last years Network Theory approaches have been successfully used for geophysical systems in the context of climate networks in which the connections among the different locations represent statistical relationships between climatic time series from these locations, inferred from correlations and other statistical methods.

In this thesis we propose a new paradigm linking the network formalism with transport and mixing phenomena in geophysical flows.

We analyze directly the network describing the material fluid flow among different locations, which we call *flow network*. Among other characteristics this network is directed, weighted, spatially embedded and time-dependent. We illustrate the general ideas with an exemplary network derived from a realistic simulation of the surface flow in the Mediterranean sea. We use network-theory tools to analyze them and gain insights into transport processes from a general point of view. We quantitatively relate dispersion and mixing characteristics, classically quantified by Lyapunov exponents, to the degree of the network nodes. A family of network entropies is defined from the network adjacency matrix, and related to the statistics of stretching in the fluid, in particular to the Lyapunov exponent field. We use a network community detection algorithm, Infomap, to partition the network into coherent regions, i.e. areas internally well mixed, but with little fluid interchange between them.

We find interesting applications of this approach to marine biology of the Mediterranean Sea. Oceanic dispersal and connectivity have been identified indeed as crucial factors for structuring marine populations and designing Marine Protected Areas (MPAs). Larvae of different pelagic durations and seasons could be modeled as passive tracers advected in a simulated oceanic surface flow from which a flow network is constructed. By applying the Infomap algorithm we extract hydrodynamical provinces from the network that result to be delimited by frontiers which match multi-scale oceanographic features. By examining the repeated occurrence of such boundaries, we identify the spatial scales and geographic structures that would control larval dispersal across the entire seascape. Based on these hydrodynamical units, we study novel connectivity metrics for existing MPAs. We also define node-by-node proxies measuring local larval retention and exchange. From the analysis of such measures we confirm that retention processes are favored along the coastlines while they are weak in the open ocean due to specific oceanographic conditions. Although these proxies were often studied separately in the literature, we demonstrated that they are inter-related under certain conditions and that their integrated analysis leads to a better understanding of metapopulation dynamics, informing both genetic and demographic connectivities.

We also consider paths in weighted and directed temporal networks, introducing tools to compute sets of paths of high probability. We quantify the relative importance of the most probable path between two nodes with respect to the whole set of paths, and to a subset of highly probable paths which incorporate most of the connection probability. These concepts are used to provide alternative definitions of betweenness centrality. We apply these tools to the temporal flow network describing surface currents in the Mediterranean sea. Despite the full transport dynamics is described by a very large number of paths we find that, for realistic time scales, only a very small subset of high probability paths (or even a single most probable one) is enough to characterize global connectivity properties of the network.

Finally we apply the same analysis to the atmospheric blocking of eastern Europe and western Russia in summer 2010. We compute the most probable paths followed by fluid particles which reveal the *Omega-block* skeleton of the event. A hierarchy of sets of highly probable paths is introduced to describe transport pathways when the most probable path alone is not representative enough. These sets of paths have the shape of narrow coherent tubes flowing close to the most probable one. Thus, as for the case of Mediterranean Sea, even when the most probable path is not very significant in terms of its probability, it still identifies the geometry of the transport pathways.

Resumen

En los últimos años ha habido grandes avances en la descripción Lagrangiana de los fenómenos de transporte y mezcla en flujos de fluidos, impulsados fundamentalmente por conceptos y métodos propios de la teoría de los sistemas dinámicos. Algunas de estas técnicas se centran en objetos geométricos - líneas, superficies - que separan regiones de fluidos con propiedades diferentes, mientras que otros se basan en el cálculo de ritmos de separación de partículas de fluido, tales como los Exponentes de Lyapunov. Por último, hay una línea de investigación que se centra en el transporte de regiones amplias de fluido, los llamados métodos "set-oriented".

Por otra parte muchos sistemas del mundo real se estudian utilizando la teoría de redes, y en los últimos años algunos sistemas geofísicos se están estudiando con éxito usando este enfoque. Es el caso de las redes climáticas, en las que las conexiones entre los diferentes lugares geograficos representan relaciones estadísticas entre series de tiempo climáticas, inferidas de correlaciones y otros métodos estadístico.

En esta tesis se propone un nuevo paradigma que une el formalismo de las redes con fenómenos de transporte y mezcla en flujos geofísicos.

Analizamos directamente la red que describe el flujo de material entre diferentes ubicaciones, que llamamos *red de flujo*. Entre otras características, esta red es dirigida, pesada, definida espacialmente y dependiente del tiempo. Presentamos las ideas generales con el ejemplo de una red derivada a partir de una simulación realista del flujo superficial en el mar Mediterráneo. Utilizamos herramientas de teoría de redes para analizarla y hacernos una idea de los procesos de transporte desde un punto de vista general. Relacionamos cuantitativamente dispersión y mezcla, clásicamente cuantificados por exponentes de Lyapunov, con el grado de los nodos de red. Una familia de entropías de red se define a partir de la matriz de adyacencia de esta, y se relaciona con los exponentes de Lyapunov. Usamos un algoritmo de detección de comunidades, Infomap, para particionar la red en regiones coherentes, es decir, áreas internamente bien mezcladas, pero con poco intercambio de fluido entre ellas.

Encontramos aplicaciones interesantes de este enfoque en la biología marina del mar Mediterráneo. De hecho la dispersión de los Océanos y la conectividad son factores cruciales para la estructuración de las poblaciones marinas y el diseño de áreas marinas protegidas. Las larvas son modeladas como trazadores pasivos trasportados por las corrientes oceánicas a partir de lo cual se construye una red de flujo. Aplicando el algoritmo Infomap extraemos provincias hidrodinámicas que resultan estar delimitadas por fronteras que coinciden con características oceanográficas de diferentes escalas. Examinando la aparición repetida de tales fronteras, identificamos las escalas espaciales y las estructuras geográficas que controlarían la dispersión de las larvas a través del entorno marino. A partir de estas unidades hidrodinámicas, estudiamos nuevos indicadores de conectividad para las áreas protegidas existentes. También definimos para cada nodo medidas de retención local de larvas y de intercambio. Del análisis de estas medidas confirmamos que se favorecen los procesos de retencióna lo largo de las costas, mientras que son irrelevantes en océano abierto. Aunque estas medidas han sido estudiadas a menudo por separado en la literatura, hemos demostrado que están, bajo ciertas condiciones, relacionadas y que su análisis integrado conduce a una mejor comprensión de la dinámica de la meta-poblaciónes, proporcionando información sobre conectividad tanto genética que demográfica.

También estudiamos caminos en redes temporales ponderadas y dirigidas, para lo cual presentamos herramientas novedosas que permiten calcular conjuntos de caminos con alta probabilidad. Cuantificamos la importancia relativa de la trayectoria más probable entre dos nodos con respecto a todo el conjunto de caminos, así como para un subconjunto de caminos altamente probable, los cuales incorporan la mayor parte de la probabilidad de conexión. Estos conceptos se utilizan para proporcionar definiciones alternativas de betweenness centrality. De nuevo aplicamos estas herramientas a la red de flujo de transporte de la superficie del mar Mediterráneo. Encontramos que , para escalas de tiempo realistas, sólo un pequeño subconjunto de caminos de alta probabilidad (o incluso el único más probable) es suficiente para caracterizar la conectividad global de la red.

Finalmente aplicamos el mismo análisis al bloqueo atmosférico de Europa del este y el oeste de Rusia que se produjo en el verano de 2010. Calculamos los caminos más probables seguidos por partículas de fluido que revelan el esqueleto del tipo *Omega-block* del evento. Una jerarquía de conjuntos de rutas altamente probables es necesario para describir las vías de transporte cuando el camino más probable por sí mismo no es suficientemente representativo. Estos conjuntos de caminos tienen la forma de tubos coherentes y estrechos que fluyen cerca del camino más probable. Por lo tanto, como para el caso de Mar Mediterráneo, incluso cuando el camino más probable no es muy significativo en términos de probabilidad, todavía identifica la geometría de las vías de transporte.

List of Publications

The contents of this thesis are published in the following articles:

- V Rossi, E Ser-Giacomi, C Lopez, E Hernandez-Garcia
  Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves
  Geophysical Research Letters **41**, 2883-2891(2014)
- E Ser-Giacomi, V Rossi, C Lopez, E Hernandez-Garcia
  Flow networks: A characterization of geophysical fluid transport
  Chaos **25**, 036404 (2015).
- E Ser-Giacomi, R Vasile, C Lopez, E Hernandez-Garcia
  Most probable paths in temporal weighted networks: An application to ocean transport
  Physical Review E **92**, 012818 (2015)
- E Ser-Giacomi, R Vasile, I Recuerda, E Hernandez-Garcia, C Lopez
  Dominant transport pathways in an atmospheric blocking event
  Chaos **25**, 087413 (2015).
- M Dubois, V Rossi, E Ser-Giacomi, S Arnaud-Haond, C Lopez, E Hernandez-Garcia
  Linking basin-scale connectivity, oceanography and population dynamics for the conservation and management of marine ecosystems.
  Under review, Global Ecology and Biogeography (2015).

Contents

Abstract iv

Resumen vi

List of Publications vii

1 Introduction 1
- 1.1 Lagrangian transport in fluid flows . . . 1
  - 1.1.1 Two points of view: Eulerian and Lagrangian . . . 1
  - 1.1.2 Dynamical systems approach . . . 2
  - 1.1.3 Lyapunov Exponents . . . 3
  - 1.1.4 Finite Time Lyapunov Exponents . . . 4
  - 1.1.5 Finite Size Lyapunov Exponents . . . 5
  - 1.1.6 Lagrangian Coherent Structures (LCS) . . . 6
  - 1.1.7 Set-Oriented Methods . . . 7
- 1.2 Complex Networks . . . 9
  - 1.2.1 Definition and basic properties . . . 9
  - 1.2.2 Community detection . . . 10
  - 1.2.3 The Infomap algorithm . . . 11
- 1.3 Networks in geophysics . . . 13
  - 1.3.1 The Climate system . . . 13
  - 1.3.2 Climate networks . . . 14
  - 1.3.3 Flow networks . . . 16
- 1.4 Plan of the thesis . . . 17

I THE FLOW NETWORK PARADIGM 19

2 Flow network construction 21
- 2.1 Discretization and Lagrangian trajectories . . . 21
- 2.2 Adjacency matrix construction . . . 22
- 2.3 Mediterranean Sea setup . . . 24

3 Local network properties 27
- 3.1 Dispersion and mixing . . . 27
- 3.2 Network entropies . . . 30

4 Community detection 35
- 4.1 Coherent regions as network communities . . . 35
- 4.2 Quality parameters . . . 36
- 4.3 Communities in the Mediterranean surface flow . . . 39

4.4 Average community descriptions 39
4.5 Comparison with spectral partitioning 41
4.6 Conclusions 45

Appendices **47**

A Relationship between network entropies and stretching statistics **47**

B Computing network entropies with nodes of differing area **51**

II ECOLOGICAL APPLICATIONS **53**

5 Marine Protected Areas (MPAs) and oceanic ecological connectivity **55**
5.1 Introduction 55
5.2 Lagrangian bio-physical modeling for the Mediterranean 58
5.3 Oceanographic characterization of hydrodynamic provinces 58
5.4 Implications for the design of MPAs 61
5.5 Conclusions 64

6 Local network measures and population dynamics **65**
6.1 Characterizing metapopulations and subpopulations in marine seascapes 65
6.2 Defining local network measures as proxies of larval connectivity 66
6.3 Hydrodynamical and genetic connectivity for management purposes . . 74
6.4 Conclusions 77

III OPTIMAL PATHS IN TEMPORAL FLOW NETWORKS **79**

7 Characterizing paths in temporal weighted networks **81**
7.1 Posing the problem: transport and probability 81
7.2 Most Probable Paths (MPPs) calculation 83
7.3 Sets of highly probable paths 86

8 Advection corridors in the Mediterranean surface flow **89**
8.1 MPPs in the ocean 89
8.2 MPP-betweenness centrality and multistep approach 91
8.3 Comparing fastest and most probable paths 93
8.4 Conclusions 95

9 Transport pathways in an atmospheric blocking event **97**
9.1 Introduction 97
9.2 Atmospheric setup 98
9.3 MPPs in the atmosphere 100
9.4 Relevance of MPPs 103
9.5 Conclusions 105

Conclusions and Perspectives **109**

Acknowledgments **111**

BIBLIOGRAPHY **111**

CURRICULUM VITAE **123**

CHAPTER 1

Introduction

1.1

Lagrangian transport in fluid flows

1.1.1 Two points of view: Eulerian and Lagrangian

The description of fluid motion can be addressed following two different ways. One can deal at any time with velocity at any spatial point in the fluid, or either, one studies the trajectory of each fluid particle. The first approach is usually called *Eulerian* and the second one *Lagrangian*. We can identify *streamlines* as lines tangent to the velocity field in each point or *trajectories* as paths followed by ideal fluid particles. The first represents a geometrical representation of the Eulerian point of view, the second of the Lagrangian one and only for the time-independent case these two objects will coincide (Vallis, 2006).

The link between these two descriptions is provided by:

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{v}(\mathbf{x}, t)\,, \\ \mathbf{x}(t_0) &= \mathbf{x}_0\,. \end{aligned} \tag{1.1}$$

In these equations, t represents time and it is the independent variable, $\mathbf{x}(t)$ represents the evolution of the initial condition x_0 at time t, i.e. the Lagrangian trajectory. The Eulerian vector function $\mathbf{v}(\mathbf{x}, t)$ typically satisfies some level of continuity.

If a fluid is characterized by a velocity field $\mathbf{v}(\mathbf{x}, t)$ and also another property ψ the material (or advective) derivative provides another relationship among the Eulerian and the Lagrangian point of view:

$$\frac{d\psi}{dt} = \frac{\partial \psi}{\partial t} + \mathbf{v} \cdot \nabla \psi\,. \tag{1.2}$$

Thinking of a fluid particle as an infinitesimal, indivisible, element of fluid the material derivative corresponds to the rate of change of a quantity ψ along the trajectory of the particle.

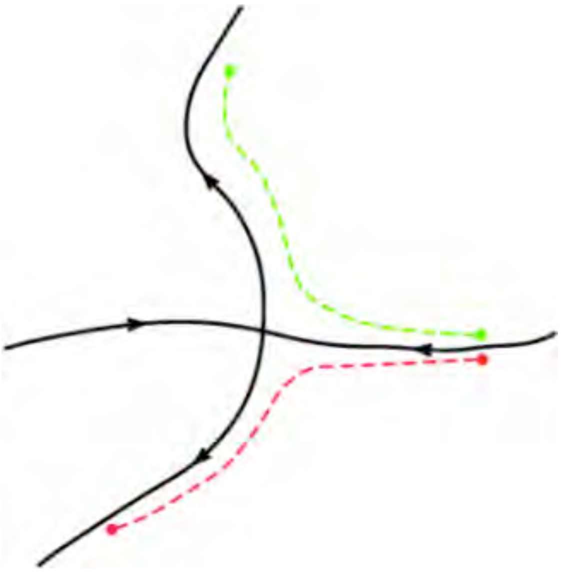

Figure 1.1: Stable and unstable manifolds (black continuous lines) intersecting at one hyperbolic point. Trajectories diverging along the stable manifold (colored dashed lines).

1.1.2 Dynamical systems approach

We can make a link between the Lagrangian description of fluid flows and dynamical systems theory by interpreting the trajectories as the evolution in time of the state of a system. Such dynamical system can be expressed by Eq. (1.1) where $\mathbf{x}(t)$ will represent the state of the system at time t. As time evolves, solutions of Eq. (1.1) trace out curves, or in dynamical systems terminology, they flow along their trajectory. Unless $\mathbf{v}(\mathbf{x}, t)$ is a linear function of the state $\mathbf{x}$ and independent of time t, and a few other cases, there is no general way to determine the analytic solution of Eq. (1.1).

When $\mathbf{v}$ is independent of time t the system is known as time-independent, or autonomous, and there are some standard techniques for studying it. We can understand the global flow geometry of time-independent systems by studying invariant manifolds of the fixed points of Eq. (1.1). In particular *stable* and *unstable manifolds* often play the most important role since they form the skeleton of motion. These concepts are described in the following. A fixed point of $\mathbf{v}$ is a point $\mathbf{x}_f$ such that $\mathbf{v}(\mathbf{x}_f) = 0$. The stable manifolds of a fixed point are all trajectories which reach the fixed point when $t \to \infty$. In a similar way, the unstable manifolds of the fixed point are all trajectories which reach the fixed point when $t \to -\infty$. The term invariant manifold is the mathematical notion of material curve. This means that the trajectory of any condition starting on the manifold, must remain on the curve. Often, stable and unstable manifolds act as separatrices, which separate regions of different motion (Hernandez-Carrasco, 2013; Bettencourt, 2014).

Let us consider now the case of a stable and an unstable manifolds intersecting at one fixed point, this point is called *hyperbolic fixed point*. In a system of differential equations a stationary hyperbolic point is a point for which the eigenvalues of the linearized system have non-zero real parts. We will see later that hyperbolicity plays an important role in transport. Fig. 1.1 shows a generic hyperbolic point and their associated stable and unstable manifolds. If we integrate the trajectories of two points initiated on both sides of a stable manifold, these points will diverge from each other forward in time. In a similar way, if we initiated two points on either sides of an unstable manifold, then these points will diverge from each other backward in time. This is the reason why these manifolds are called separatrices, since they separate qualitatively different trajectories.

Considering now time-dependent dynamical systems, we find that they also have regions of dynamically distinct behavior which can be thought of as being divided by separatrices. However, for such systems these regions change over time, and hence so do the separatrices, and one should not read into the analogy between these separatrices and traditional definitions of stable and unstable manifolds too much. These are the so called *Lagrangian Coherent Structures* (LCS) and are *material curves*, which means that represent curves that cannot be crossed by other fluid particle trajectories.

Nowadays the techniques used to detect these structures can be roughly classified as follows. On the one hand, some approaches focus on the geometric objects - lines, surfaces - separating fluid regions with different properties that are identified with LCS (Mancho et al., 2003, 2004). On the other hand, another class of algorithms have focussed on computing stretching-like fields in the fluid domain, such as different types of Lyapunov exponents or other Lagrangian descriptors (Haller, 2001; Shadden et al., 2005; Aurell et al., 1997; Lacorata et al., 2001). Ridges or singular lines in such fields turn out to be related, under suitable conditions, to the LCS too. Finally, there is a line of research focussing on the moving fluid regions themselves, the so-called set-oriented methods (Dellnitz et al., 2001; Froyland and Dellnitz, 2003). The geometric approaches are designed to follow specific structures during particular transport events, whereas the coarse-graining inherent to the set-oriented methods makes them useful also to estimate statistical properties in more extended space and time intervals. Stretching-field methods can be used to follow particular events or, by simple averaging, also to characterize dispersion and stirring statistics in large areas or long times.

1.1.3 Lyapunov Exponents

In 1963, Edward Lorenz described the behavior related to chaotic attractors occurring in dissipative systems (Lorenz, 1963), and recognized the unpredictability of chaotic behavior in connection with the numerical solution of an atmospheric model. Chaotic motion is irregular in time, unpredictable in the long term and sensitive to initial conditions, and complex, but ordered, in the phase space. In present day literature a system with bounded trajectories is said to be chaotic if small i.e. infinitesimal perturbations grow exponentially with time, and this is connected to a positive Lyapunov exponent.

The classical or global *Lyapunov exponent* is defined as the exponential rate of separation, averaged over infinite time, of particle trajectories initially separated infinitesimally. Consider $\mathbf{x}(t_0)$ and $\mathbf{x}(t) = \mathbf{x}(t_0) + \delta\mathbf{x}(t)$ as two particle trajectories separated initially by a distance $\delta\mathbf{x}(t_0)$. The global Lyapunov exponent is defined by:

$$\lambda = \lim_{t\to\infty}\left(\lim_{\delta\mathbf{x}(t_0)\to 0}\frac{1}{t}\log\frac{|\delta\mathbf{x}(t)|}{|\delta\mathbf{x}(t_0)|}\right). \tag{1.3}$$

Negative Lyapunov exponents are characteristic of stable solutions of dissipative or non-conservative systems (the damped harmonic oscillator for instance). Such systems

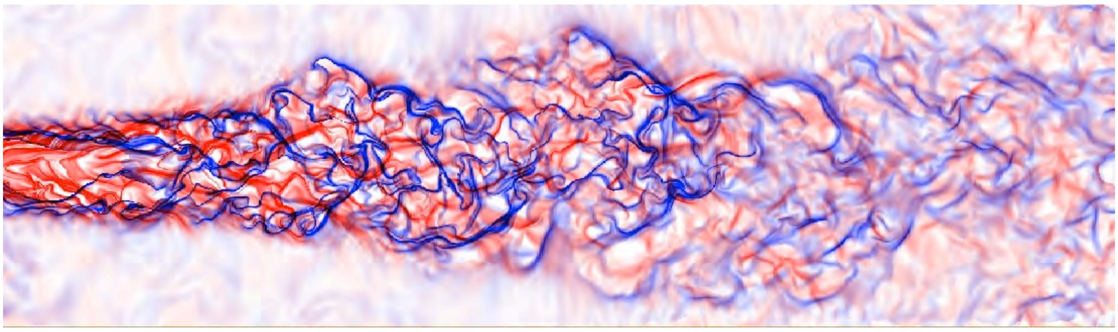

Figure 1.2: FTLE snapshot for an unsteady ideal flow. Forward (red) and Backward (blue) in time FTLE fields. (Figure from Comparative Visualization Group-ZIB)

exhibit asymptotic stability (the orbit attracts to a stable fixed point). If λ is positive, the orbit is unstable and chaotic. Nearby points will diverge. For a continuous system, the phase space would be a tangled sea of wavy lines like a pot of spaghetti. Whereas the (global) Lyapunov exponent gives a measure for the total predictability of a system, it is sometimes interesting to estimate the local predictability around a point $\mathbf{x}$ in phase space. In this case, a generalization of the Lyapunov exponent, called the *local Lyapunov exponent* (LLE), has been proposed to study the growth of non-infinitesimal perturbations in dynamical systems. Recently the concept of a LLE has been applied to study dispersion in turbulent flow fields (Barreira and Pesin, 2002). The LLE is a scalar value which characterizes the amount of stretching about the trajectory of point $\mathbf{x}$ over a time interval. For most flows of practical importance, the LLE varies as a function of space and time. The LLE is not an instantaneous separation rate, but rather measures the average, or integrated, separation between trajectories. This distinction is important because in time-dependent flows, the instantaneous velocity field often is not very revealing about actual trajectories, that is, instantaneous streamlines can quickly diverge from actual particle trajectories. However the LLE accounts for the integrated effect of the flow because it is derived from particle trajectories, and thus is more indicative of the actual transport behavior. There are two types of LLE: FTLE and FSLE, distinguished in the following.

1.1.4 Finite Time Lyapunov Exponents

Dealing with time-dependent flows taking a limit $t \to \infty$ in Eq. (1.3) has no meaning anymore. If we want to define a Lyapunov exponent for the time dependent case we have to eliminate this asymptotic character. We define thus the *Finite Time Lyapunov Exponent* (FTLE) as:

$$\lambda(\mathbf{x}_0, t_0, \tau) = \lim_{\delta\mathbf{x}(t_0)\to 0} \frac{1}{|\tau|} \log \frac{|\delta\mathbf{x}(t_0+\tau)|}{|\delta\mathbf{x}(t_0)|} \,. \tag{1.4}$$

The FTLE is a function of the initial position $\mathbf{x}$ at time t_0, but if we vary t_0, then it is also a function of time (see Fig. 1.2 for a graphical representation of a FTLE field). We assume $\delta\mathbf{x}(t_0)$ is infinitesimal and the orientation is chosen so that λ is maximal. To deal conveniently with the infinitesimal condition, one can obtain FTLEs using the flow map $\phi_{t_0}^{\tau}$, defined as the function that takes the initial position $\mathbf{x}_0$ at time t_0 of any fluid element advected by the flow to its final position $\phi_{t_0}^{\tau}(\mathbf{x}_0)$. If we consider a small perturbation q to the initial condition we have:

$$\phi_{t_0}^{\tau}(\mathbf{x}_0 + \mathbf{q}) = \phi_{t_0}^{\tau}(\mathbf{x}_0) + \frac{\partial \phi_{t_0}^{\tau}}{\partial \mathbf{x}_0}\mathbf{q} + O(||\mathbf{q}||^2)\,. \tag{1.5}$$

From this approximation, and assuming that the $O(||\mathbf{q}||^2)$ term is negligible, we have that:

$$\delta\mathbf{x}(t_0 + \tau) = \phi_{t_0}^{\tau}(\mathbf{x}_0 + \mathbf{q}) - \phi_{t_0}^{\tau}(\mathbf{x}_0) \approx \frac{\partial \phi_{t_0}^{\tau}}{\partial \mathbf{x}_0}\mathbf{q} = \nabla\phi_{t_0}^{\tau}(\mathbf{x}_0)\mathbf{q}\,, \tag{1.6}$$

and the length of the perturbation after the integration time τ is given by the norm (using the inner product):

$$||\delta\mathbf{x}(t_0 + \tau)|| = \sqrt{\langle \nabla\phi_{t_0}^{\tau}(\mathbf{x}_0)\mathbf{q}, \nabla\phi_{t_0}^{\tau}(\mathbf{x}_0)\mathbf{q}\rangle} = \sqrt{\langle \mathbf{q}, C_{t_0}^{\tau}(\mathbf{x_0})\mathbf{q}\rangle}\,, \tag{1.7}$$

where $C_{t_0}^{\tau}(\mathbf{x_0})$ is the finite-time version of the Cauchy-Green deformation tensor defined by (Haller, 2001; Shadden et al., 2005) :

$$C_{t_0}^{\tau}(\mathbf{x_0}) = (\nabla\phi_{t_0}^{\tau}(\mathbf{x}_0))^T(\nabla\phi_{t_0}^{\tau}(\mathbf{x}_0))\,. \tag{1.8}$$

In the case of maximum stretching of the perturbation, the vector $\delta\mathbf{x}(t_0)$ is aligned with the eigenvector associated with the maximum eigenvalue of $C_{t_0}^{\tau}(\mathbf{x_0})$. If $\sigma(\mathbf{x}_0, t_0, \tau)$ is such largest eigenvalue, then the value of the FTLE at the point $\mathbf{x_0}$ at time t_0 and depending on the integration time τ is:

$$\lambda(\mathbf{x}_0, t_0, \tau) = \frac{1}{|\tau|}\log\sqrt{\sigma(\mathbf{x}_0, t_0, \tau)}\,. \tag{1.9}$$

1.1.5 Finite Size Lyapunov Exponents

If we want to know instead about the predictability time with respect to a finite perturbation, it should be determined by a quantity analogous to the Lyapunov exponent. The natural starting point is the time it takes for a perturbation to grow from an initial size δ to a given but arbitrary threshold value Δ. This is called the (δ, Δ)-predictability-time and denoted by $T(\delta, \Delta)$ (Aurell et al., 1997).

We define the *Finite Size Lyapunov Exponent* (FSLE) as an average of some function of the predictability time, such that if both δ and Δ are in the infinitesimal range, we will recover the usual Lyapunov exponent. An obvious choice is then:

$$\lambda(\delta, \Delta) = \frac{1}{\langle T(\delta, \Delta)\rangle}\log\left(\frac{\Delta}{\delta}\right), \tag{1.10}$$

where the average is over different initial conditions. In contrast to infinitesimal perturbations, for finite perturbations the threshold Δ is typically not to be taken much larger than the finite perturbation δ. In contrast with FTLE now we have relaxed the limit $\delta\mathbf{x}(t_0) \to 0$. FSLE can be used to study dispersion of particles in flows. In this case the perturbations are distances between particles, and the given threshold of tolerance Δ can be defined to be proportional to the initial perturbation δ, and it is standard to define FSLE in the following way (Lacorata et al., 2001),

$$\lambda(\delta) = \frac{\log(r)}{\langle \tau(\delta, r\delta)\rangle}\,, \tag{1.11}$$

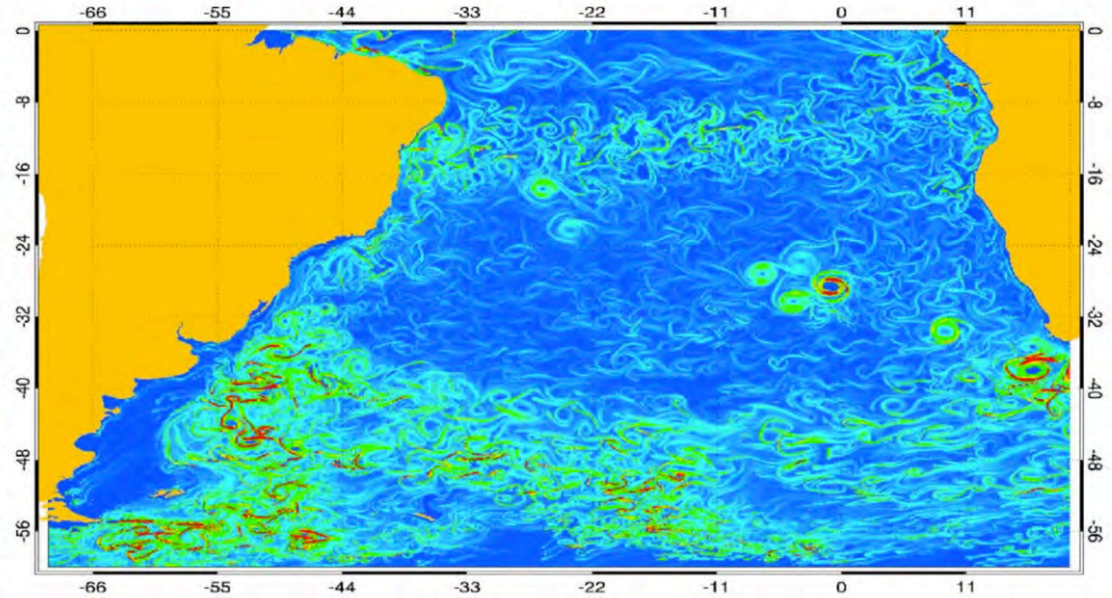

Figure 1.3: A snapshot of the FSLE field in the Atlantic Ocean. (Figure from (Hernández-Carrasco et al., 2012))

where $\langle \tau(\delta, r\delta) \rangle$ is the average time (over the number of particles pairs) required to separate from a distance of δ to $r\delta$.

We can now define a local version of the FSLE following the same approach used for the FTLE (d'Ovidio et al., 2004; Hernández-Carrasco et al., 2012, 2011; Bettencourt, 2014; Hernandez-Carrasco, 2013). Focusing on the point $\mathbf{x}_0$ we can define the associated local FSLE as:

$$\lambda(\mathbf{x}_0, t_0, \delta_0, \delta_f) = \frac{1}{|\tau|} \log \frac{\delta_f}{\delta_0} . \tag{1.12}$$

where δ_i and δ_f are now fixed parameters and τ is the variable that will determine the FSLE at each point. It corresponds to the time needed for two initial conditions separated by an initial distance δ_0 at time t_0 to reach a final separation δ_f (see Fig. 1.3 for a graphical representation of the FSLE field in the ocean).

1.1.6 Lagrangian Coherent Structures (LCS)

In time-dependent turbulent flows, coherent structures tend to emerge. These are dynamic regions where a flow quantity exhibits a high degree of correlation. In 2d flows, coherent structures are identified with vortices embedded in a background of weak turbulence. In between these vortices, thin filaments of concentrated vorticity exist. From the Lagrangian point of view, visualizations of densely populated patches of passive particles in 2d flows lead to the observation of distinct behaviors: continued elongation and thinning of initially thick blobs of particles; coherent rotation of particle patches or simple, correlated, translational motions of particles. From observations of these kinds of coherent motions, distinct LCS can be educed:

- Local stretching in the flow appears to happen across LCS boundaries through divergence of nearby particles.

- Thinning and folding appears to happen along LCS boundaries, as particles follow the curved boundaries

The first behavior is related to instability of particle trajectories due to which divergence occurs while the second resembles the behavior of particle trajectories converging towards an unstable manifold of the flow. For this reason LCS can be seen as material lines that create smaller tracer scales through stability or instability (Haller and Yuan, 2000). Analytic criteria for the identification of such structures were derived for flows in 2d in (Haller and Yuan, 2000; Haller, 2000). These criteria were based firstly on the computation of the time interval during which a material surface would be stable or unstable, coherent structures being those that maximized this measure (the hyperbolicity time approach). Another alternative method put forth was the identification of coherent structures with the local extrema of the FTLE field. The rationale for this criteria is that if coherent structures attract or repel particles for the longest in the flow, then, the net growth of a disturbance normal to them should be the largest during the time interval of interest (Haller, 2001). A third common approach is to use instead FSLEs, which are a measure of the separation rate of fluid particles between two given distance thresholds to identify LCS. Although a rigorous connection between the FSLE and LCS has not been established, several works have shown that the ridges of the FSLE behave in a similar fashion as the ridges of the FTLE field (d'Ovidio et al., 2004; Joseph and Legras, 2002; Branicki and Wiggins, 2010). Note that there are further methods serving similar purposes like, for instance, the leaking approach (Schneider and Tél, 2003; Schneider et al., 2005) but it is not the purpose of this thesis to provide a complete overview of them.

1.1.7 Set-Oriented Methods

Until now we focused on the study of single trajectories or of the local rate of separation of pairs of initial conditions. We have seen that this information is particularly helpful in the investigations of geometric properties of a dynamical system. However there is also important set-related information covering both topological and statistical aspects of the underlying dynamical behavior, especially for the case of complicated dynamics. With the aim to capture the global structure of a given dynamical system, we can set up our analysis by using global set-oriented methods rather by an approach based on long term computations of single trajectories. Objects of interest become thus *invariant sets* and *almost-invariant sets*: collections of sets that behave in a very predictable way, in spite of the fact that individual trajectories are entirely unpredictable (see Fig. 1.4 for a three-dimensional representation of almost-invariant sets in a simple ideal flow). This approach is based on a discretization of the dynamical system under study that allows to introduce a probabilistic point of view and at the same time to use tools from graph theory.

In a time independent system we can consider $T : X \to X$ to be a continuous mapping that defines a discrete time dynamical system on its chain recurrent set X. We think of an almost-invariant set as a set $A \subset X$ such that $T(A)$ is not very different from A (Dellnitz et al., 2001; Froyland and Dellnitz, 2003; Froyland, 2005; Froyland and Padberg, 2009). A way to quantify this is to consider the fraction of Lebesgue measure m, that stays

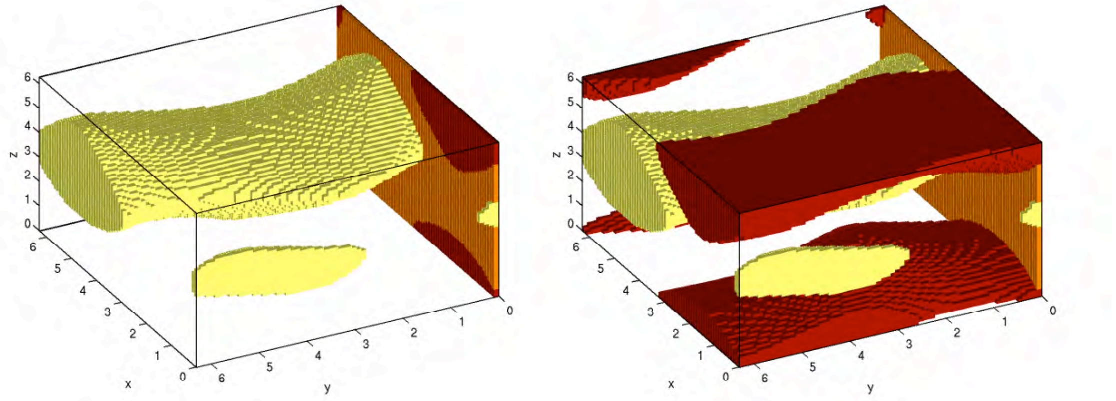

Figure 1.4: Three-dimensional almost invariant sets extraction from a theoretical flow. Different colors correspond to different almost invariant sets (Figure from (Froyland and Padberg, 2009))

within A. We do this defining the ratio:

$$\rho(A) = \frac{m(A \cap T^{-1}A)}{m(A)} \,. \tag{1.13}$$

With the goal of identifying regions of qualitatively different dynamical behavior we seek a partition of the system in q almost-invariant sets $A_1, ..., A_q$ such that the $\rho(A_k)$ are all close to one. This corresponds to maximizing the quantity:

$$\rho(A_1, ..., A_q) = \frac{1}{q} \sum_{k=1}^{q} \rho(A_k) \,. \tag{1.14}$$

Finding measurable sets $A_1, ..., A_q$ maximizing ρ is an infinite-dimensional optimization problem. We reduce this to a finite-dimensional problem by creating a fine box partition $\mathcal{B} = \{B_1, ..., B_n\}$ of a covering of the chain recurrent set T. Switching to this discretized representation allow us to define a weighted transition matrix for our dynamical system. We think of discretizing the smooth dynamics to form a finite state Markov chain with transition matrix P_{ij} given by:

$$P_{ij} = \frac{m(B_i \cap T^{-1}B_j)}{m(B_i)} \,. \tag{1.15}$$

We note that the matrix element P_{ij} can now be interpreted as a transition probability from B_i to B_j. Moreover the matrix **P** has a graph representation, where the nodes of the graph correspond to states of Markov chain. When $q = 2$ our optimization problem can be translated to the problem of finding a minimal cut for this graph. It is clear now the link between graph theory and dynamical systems and many algorithms can be applied to approximate the minimal cut solution.

When dealing instead with time dependent systems it is necessary to introduce the concept of coherence. Sets indeed are now transported and deformed by the flow and we aim to understand how much they remain coherent and nondispersive under the temporal evolution. These regions, called *coherent sets*, are robust to perturbation and are carried along by the chaotic flow with little transport between the coherent sets and the rest of state space (Froyland et al., 2010; Santitissadeekorn et al., 2010).

1.2

Complex Networks

1.2.1 Definition and basic properties

A graph is essentially a way to code a relation (physical links, interactions etc.) between the elements of a system. The elements of the system identify the set V (set of vertices), and the relations among those the set E (set of edges). Complex networks are graphs which possess non-trivial topological features, as patterns of connections between their elements that are neither purely regular nor fully random. With such a level of generality it is easy to understand that a wide array of systems can be studied within the network theory perspective (Caldarelli, 2007; Newman, 2009).

The graph indicated as $G(V, E)$ can be drawn plotting the vertices as points and the edges as lines between them. It is not important how they are actually drawn. Ultimately the only thing that matters is to know which vertices are connected. Whenever a real number can be attached to an existing edge between a pair of nodes (i, j) we have that the edge is characterized by a weight w_{ij}. The graph in this case is a *weighted* graph. If w_{ij} could be different from w_{ji} we say that the graph is also *directed* i.e. links can be characterized by a direction. Approaching many real systems we need to deal with their intrinsic time-variability, from the network perspective this is translated in studying a graph that is varying its connections in time. Thus, after a proper choice of the time resolution, we can look at several snapshots of the graph at different moments as parts of a time-dependent network describing the system over the whole time window considered. We say in this case that we are dealing with a *temporal* network (Holme and Saramäki, 2012).

A complete description of a graph of N vertices is provided by its $N \times N$ adjacency/weight matrix A. If the graph is unweighted each element A_{ij} can be either 0 or 1 where 1 means a link among nodes i and j and 0 no link. While if links are weighted we will simply have $A_{ij} = w_{ij}$.

From the information contained in **A** we can define measures helpful to understand the importance of single nodes across the network. The degree of a node is the number of first neighbors the node has. For the more general case of a directed network we can distinguish between *in-degree* (referring at the number of in-coming links) and *out-degree* (referring instead at the number of out-coming links) of the node i:

$$K_i^{in} = \sum_{j=1}^{N} \tilde{A}_{ji}\,, \tag{1.16}$$

$$K_i^{out} = \sum_{j=1}^{N} \tilde{A}_{ij}\,.$$

Where $\tilde{\mathbf{A}}$ is the adjacency matrix: $\tilde{A}_{kl}$ it is equal to one if $w_{kl} \neq 0$ and zero otherwise. While the degree is only the number of the connections of each node, we can take into account also of the strengths of such connections defining the *in-strength* and the *out-strength* of the node i:

$$S_i^{in} = \sum_{j=1}^{N} A_{ji}, \tag{1.17}$$
$$S_i^{out} = \sum_{j=1}^{N} A_{ij} .$$

Degree and strength represent the simpler centrality measures of nodes in the network. High values will be characteristics of hubs and/or nodes crucially important in the dynamics. Behind these first-order quantities we can define other measures such as clustering coefficient, eigenvector centrality, Page Rank, etc. (Caldarelli, 2007; Newman, 2009).

Introducing the concept of *path* we can finally define more complex measures based on path-connectivity like betweenness and closeness. A path between a pair of network nodes is defined as a set of links connected between them and connecting the start node with the final one. Among the several paths relating two nodes it is usual, for unweighted networks, to pick up the *shortest* one i.e. the path composed by fewer links (when dealing with weighted networks many choices are possible to define the length of a path). With the complete set of paths at hand we can now define the *betweenness centrality* of a node i as:

$$\mathcal{B}_i = \sum_{k,l} \frac{\sigma_{kl}^i}{\mathcal{N}_P} , \tag{1.18}$$

where σ_{kl}^i is equal to one if the shortest path connecting the pair k, l is passing across i and zero otherwise. The betweenness is normalized to the total number of paths on the network $\mathcal{N}_p$. We define also the *closeness centrality* of i as:

$$C_i = \frac{1}{\sum_k d(i,k)} , \tag{1.19}$$

where $d(i,k)$ is the length of the shortest path connecting the pair i, j.

1.2.2 Community detection

While centrality measures can inform us about the local connectivity features of the network (i.e. node-by-node linking properties), another long-standing question is about understanding the global structure of the network trying to individuate big groups of nodes with different dynamical/connectivity features. This means looking for a partition of the network in communities such that each of them results to be strongly linked internally but with few connections with the others, this is commonly known as the *community detection* problem. Finding communities within an arbitrary network can be computationally difficult. The number of communities, if any, within the network is typically unknown and the communities are often of unequal size and/or density. Despite these difficulties, however, several methods for community finding have been developed and employed with varying levels of success. A commonly used algorithm for finding communities is for instance the Girvan-Newman algorithm. This algorithm identifies edges in a network that lie between communities and then removes them, leaving behind just the communities themselves. The identification is performed by employing the graph-theoretic measure betweenness (see Eq. 1.18), which assigns a number to each edge which is large if the edge lies between many pairs of nodes (Girvan and Newman, 2002). Another standard method is, in spite of its known drawbacks, modularity maximization (Newman, 2004). Modularity is a benefit function that measures the quality of a particular division of a network into communities. The modularity maximization method detects communities by searching over possible divisions of a network that has particularly high modularity. Since exhaustive search over all possible divisions is usually intractable, practical algorithms are based on approximate optimization methods such as greedy algorithms or simulated annealing, with different approaches offering different balances between speed and accuracy.

Approaching the same problem from the point of view of Subsection 1.1.7 we see that communities can be interpreted as invariant (or coherent) sets of the dynamical system represented by the network. The optimal cut principle will lead thus to a partition of the network in communities with a minimal exchange among them. A standard way to find such partition is the spectral method (Fiedler, 1975). It consists in using the first eigenvectors (Froyland and Dellnitz, 2003; Froyland, 2005; Froyland and Padberg, 2009) (or singular vectors for the time-dependent case (Froyland et al., 2010; Santitissadeekorn et al., 2010)) of the transition matrix (or of the associated Laplacian matrix), they detect indeed regions of the system of maximal invariance . By thresholding or clustering such eigenvectors or singular vectors is possible thus to find a partition in sets of the whole system.

1.2.3 The Infomap algorithm

We will focus now on one algorithm called *Infomap* that is the one used for the studies described in this thesis. *Infomap* (Rosvall and Bergstrom, 2008) is a community-detection algorithm (Newman, 2009; Danon et al., 2005; Lancichinetti and Fortunato, 2009; Fortunato, 2010; Aldecoa and Marín, 2013) that retains both the "direction" and "weight" information of each link in the network. *Infomap* does not require to specify *a priory* the number of communities to be detected. It does not assume communities with similar sizes (as for example spectral partitioning mentioned before (Froyland, 2005; Froyland et al., 2007)) nor suffers from the 'resolution limit' (Fortunato and Barthélemy, 2007) which limits the minimum community size detectable by most algorithms. In addition to these convenient properties, the minimization algorithm is efficiently implemented in publicly available software. *Infomap* considers an ensemble of random walkers in the weighted and directed network defined by $\mathbf{A}$, moving with the transition probabilities in that matrix. Then, the method considers from the information-theory point of view the optimal coding of the ensemble of possible random walks. To this end the network is divided in communities and each random walk is coded by sequences of words that represent successive locations inside a community and jumps between different communities (see Fig. 1.5). The information-theoretic lower bound to the average length of the codeword used is given in terms of the transition probabilities

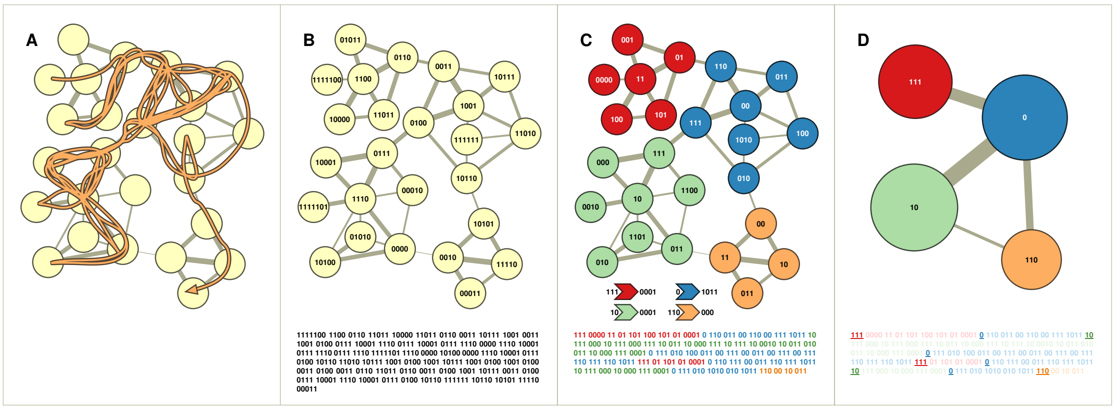

Figure 1.5: Graphical representation of the *Infomap* algorithm. (A) The orange line shows one sample trajectory of a random walk on the network. (B) The aim is to give unique names to important structures of the network. (C) It is possible to give a two-level compressed description of the random walk, in which major clusters receive unique names, but the names of nodes within clusters are reused. (D) Reporting only the module names, and not the locations within the modules, provides finally an efficient coarse graining of the network (Figure from (Rosvall and Bergstrom, 2008))

and of the specific partition in communities by the so-called *map equation*:

$$L = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{\alpha=1}^{c} p_{\circlearrowright}^{\alpha} H(\mathcal{P}^{\alpha})\,. \tag{1.20}$$

Where c is the number of communities in the particular partition considered. The first term involves the Shannon entropy associated to the transitions between different communities α:

$$H(\mathcal{Q}) = -\sum_{\alpha=1}^{c} \frac{q_{\alpha\curvearrowright}}{q_{\curvearrowright}} \log_2 \left(\frac{q_{\alpha\curvearrowright}}{q_{\curvearrowright}} \right). \tag{1.21}$$

Where $q_{\alpha\curvearrowright}$ is the probability to leave community α in one random-walk step, and $q_{\curvearrowright} = \sum_{\alpha=1}^{c} q_{\alpha\curvearrowright}$. Expressions for these quantities in terms of the components of the network matrix $\mathbf{A}$ are given in Rosvall and Bergstrom (2008). The second term in Eq. (1.20) contains the Shannon entropies $H(\mathcal{P}^{\alpha})$ associated to the words used to codify the position inside a community α and the word that denote the exit from that community:

$$H(\mathcal{P}^{\alpha}) = -\sum_{i\in\alpha} \frac{\pi_i}{p_{\circlearrowright}^{\alpha}} \log_2 \left(\frac{\pi_i}{p_{\circlearrowright}^{\alpha}} \right) - \frac{q_{\alpha\curvearrowright}}{p_{\circlearrowright}^{\alpha}} \log_2 \left(\frac{q_{\alpha\curvearrowright}}{p_{\circlearrowright}^{\alpha}} \right). \tag{1.22}$$

The notation $i \in \alpha$ indicates sum over the nodes pertaining to community α. π_i is the stationary distribution of the random walk and $p_{\circlearrowright}^{\alpha} = q_{\alpha\curvearrowright} + \sum_{i\in\alpha} \pi_i$. Again, expressions for these quantities can be obtained from the elements in the network matrix $\mathbf{A}$ (Rosvall and Bergstrom, 2008). *Infomap* finds the partition that minimizes the quantity in (1.20), i.e. the partition that provides a shorter description of the ensemble of walks going in and outside the communities. In other words, it finds the partition for which the random walks remain most of the time inside the communities with few jumps between them. This minimization process uses a deterministic greedy algorithm followed by a simulated-annealing which is repeated several times to select the best partition.

1.3

Networks in geophysics

1.3.1 The Climate system

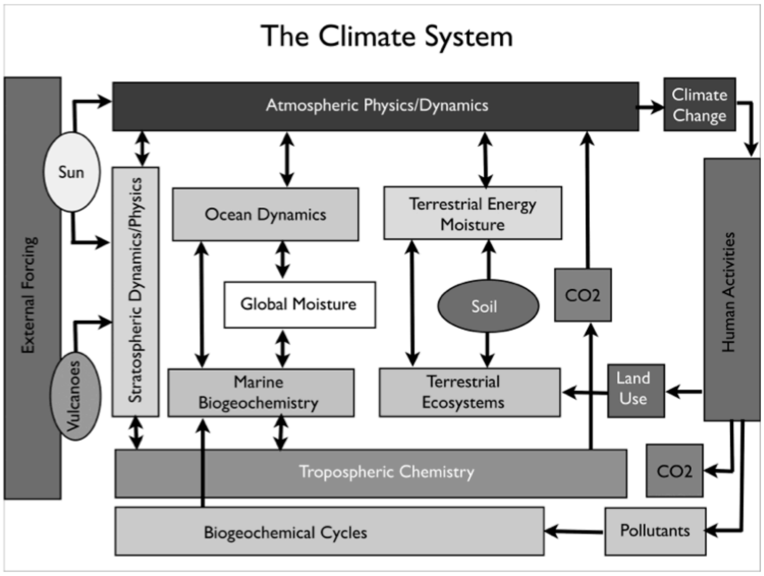

Figure 1.6: A schematic representation of the several components of the climate system and and their interactions.

The climate system is made up of building blocks which are based on elementary physical principles, but which have surprising and profound collective behavior when allowed to interact on the planetary scale. This is why we can say that the climate is a complex system. Feedbacks and interactions among its many components make this complexity hard to explain (see Fig. 1.6). Moreover climatic phenomena present a strong variability in a wide range of spatial and temporal scales. Interactions among several different scales make even more difficult to extract and isolate clear patterns from the whole dynamics.

This complexity could be partially understood explaining how climate phenomena in one geographical region affect the climate in other regions. It is widely accepted indeed that climate in a region is not the result of local factors, as the atmosphere connects far away regions through waves and advection of mass, heat and momentum. These long-range couplings are called *teleconnections* and have been shown to be an ubiquitous feature of atmospheric dynamics. In this context *climate networks* have been introduced (Tsonis and Roebber, 2004; Tsonis et al., 2006; Yamasaki et al., 2008; Donges et al., 2009a), representing a powerful tool to detect spatiotemporal patterns in the climate system (Deza, 2015).

1.3.2 Climate networks

The application of complex network theory to climate science makes possible to extract relevant information about a system without yielding an oversimplification, or being forced to handle the full-scale detailed model which can obscure the interpretations or being computationally too expensive.

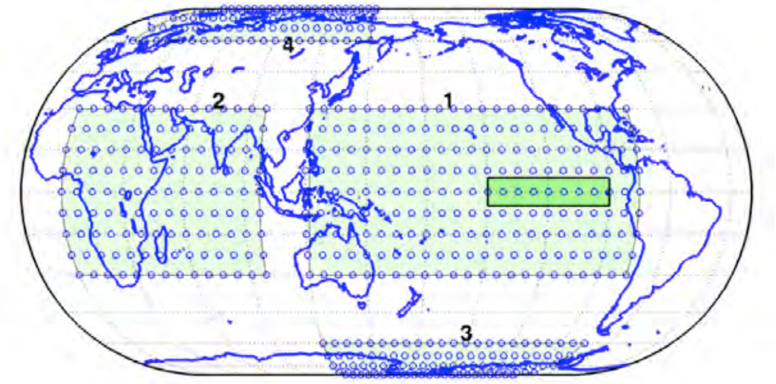

Figure 1.7: An example of grid of climate data. (Figure from (Yamasaki et al., 2008))

The vertices of a climate network are identified with the spatial grid points of an underlying global climate data set. Due to the spherical geometry of the earth, a latitude-longitude based nodes scheme will yield an inhomogeneous distribution of nodes in physical space and this is why the connectivity is weighted with the area associated to each grid point in the data set considered (see Fig. 1.7). Edges are added between pairs of vertices depending on the degree of statistical interdependence between the corresponding pairs of anomaly time series taken from the climate data set. Commonly used variables to define the data sets are, for instance, near surface air temperature (SAT) or sea surface temperature (SST). Data are typically obtained from a data assimilation system i.e. a general circulation model (GCM) employed to fill in the gaps, in space and in time, in real observed data. The analysis is usually performed using data anomalies, which are calculated as the departure from the climatological mean. This allows to extract properly the variability patterns in the data sets.

The climate network properties will depend on the methodology employed to infer the presence of connections between two nodes, i.e., the similarity measure used to include a particular link in the network and the procedure to filter out those correlations that may have occurred merely by chance. In order to construct a climate network, lets suppose there is a correlation measure and a method to assess the significance of the links. A correlation matrix C_{ij} can be calculated and after an appropriate significance test, an adjacency matrix A_{ij}, summarizing the connectivity of the network, can be finally constructed. Many different similarity measures can be employed to calculate C_{ij} and they are divided into two big families: *linear* and *non-linear* techniques. Linear analysis include Pearson cross correlation calculation. Non-linear information techniques are mostly based on information theory principles. It is possible to perform entropy calculations, which is a measure of the information content, or the lack of it, in a time series. Mutual information could also be defined as a measure of similarity between two time series, which takes into account the information shared by the two time series. It is also possible to use symbolic ordinal analysis using ordinal patterns. Whatever is

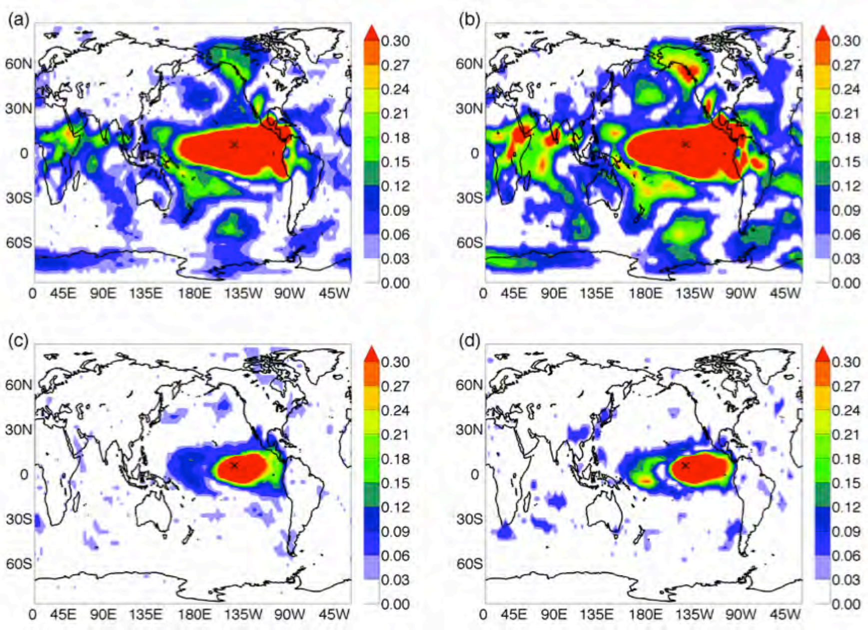

Figure 1.8: Examples of area weighted connectivity patterns of a node in central Pacific (indicated with x) with the whole climate network. (Figure from (Deza et al., 2013))

the measure that we use, it is fundamental then a study of statistical significance of the inferred connections. It consists of determining whether the outcome of the analysis would lead to a rejection based on a pre-specified low probability threshold called *p-value*. This is translated in defining a *null hypothesis* for our system. It can be done defining p-values only from observations or using *surrogates* (data not coming from observations but responding to the null hypothesis) (Deza, 2015).

After constructing such networks it is possible finally to apply many powerful tools from Network Theory to extract information from the system under study (see for instance Fig. 1.8 and 1.9). In this network approach results extremely helpful giving at the same time a robust coarse grained description of the phenomena from a complex system point of view (i.e. an interacting and non-linear perspective) (Tsonis and Roebber, 2004; Tsonis et al., 2006; Tsonis and Swanson, 2008a; Yamasaki et al., 2008, 2009; Donges et al., 2009b,a; Tsonis et al., 2010; Donges et al., 2011b). Climate networks have been constructed from temperature data for “El Niño”, “La Niña” and neutral years, and showed that during the Niño years many links were broken in comparison to the other two cases. They further suggested that the number of surviving links could be used as a measure for gauging “El Niño” events, and the stability of the system in the various ENSO phases (Yamasaki et al., 2008; Tsonis and Swanson, 2008b). A multilayer network of geopotential height (GH) at different heights was studied in (Donges et al., 2011a). A measure called “cross-betweenness” was defined in order to help to identify regions which are particularly important for mediating vertical wind field interactions as interaction between the GH networks. It was found that within the different phases of the North-Atlantic Oscillation (NAO) the correlation values of the links in a climate network covering the North Atlantic for wintertime data was significantly different. This was also proposed as a measure to track the NAO pattern (Guez et al., 2012). The

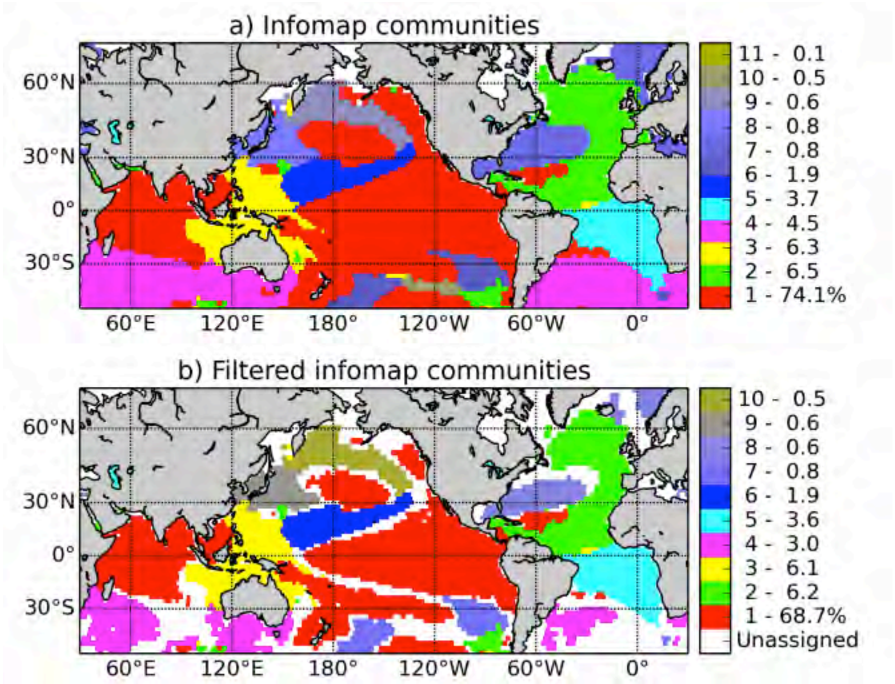

Figure 1.9: Examples of communities extracted from climate networks. (Figure from (Tantet et al., 2014))

interannual to multidecadal timescales variability patterns in sea surface temperature was also studied through the community structure of interaction networks constructed from SST. The community structure was interpreted using known dominant patterns of variability, such as the "El Niño"/Southern Oscillation and the Atlantic Multidecadal Oscillation (AMO). The study of the relationship between the communities and indices of global surface temperature showed that, while ENSO was most dominant on interannual timescales, the Indian West Pacific and North Atlantic may also play a key role on decadal timescales (Tantet et al., 2014). Another use of Climate networks has been also as predictors of regime change. New early warning indicators of the collapse of the Atlantic Meridional Overturning Circulation (MOC) were proposed (Mheen et al., 2013). These indicators are based on a climate network of spatial correlations in the time series of the Atlantic ocean temperature field. A meridional-depth model of the MOC is used for which the critical conditions for collapse can be explicitly computed. This network is used to monitor changes in spatial correlations in the model temperature time series as the critical transition is approached. The indicators were based on changes in topological properties of the network, in particular changes in the distribution functions of the degree and the clustering coefficient.

1.3.3 Flow networks

In this thesis we are interested in building networks from physical connections instead than of statistical correlations. This means to focus our analysis on mass transport driven by geophysical flows. The resulting networks will be called *flow networks*. The main goal is, after fixing spatial and time scales, to give a clear description on how transport is organized in space. We stress again that in geophysical flows turbulence and time dependance make difficult to extract clear and evident patterns.The network representation permits to overcome this problem allowing us to coarse-grain the system in order to evidence the skeleton of the transport dynamics. At the same time we are still representing the highly non-linear dynamics characterizing these kind of flows. In this thesis we will restrict our treatment to the case of 2-dimensional transport , but generalizations to 3-dimensions are not conceptually different.

1.4

Plan of the thesis

Part I is devoted to the introduction of the concept of flow networks in geophysical flows applied to the study of the surface circulation in the Mediterranean Sea. After describing the method (Chapter 2), we focus on local network properties (in particular defining a family of network entropies) and their relations with dispersion and mixing processes (Chapter 3). Finally we characterize the community structure of the networks comparing our results with the spectral partitioning method (Chapter 4) (Ser-Giacomi et al., 2015a). In Part II we propose applications of our approach to marine biology and ecology. We highlight the ecological importance of Marine Protected Areas (MPAs) for biological processes (Chapter 5) and studying the linking properties of flow networks in the Mediterranean Sea we provide novel metrics to understand population dynamics and connectivity across the whole seascape (Chapter 6) (Rossi et al., 2014; Dubois et al., 2015). In Part III we study long range connectivity patterns in flow networks by defining most probable paths (MPPs) and highly probable paths (HPPs) in weighted temporal networks (Chapter 7). By analyzing such sets of paths in networks describing water transport in the Mediterranean Sea (Chapter 8) (Ser-Giacomi et al., 2015b) and air masses transport in the Eastern Europe and Russian atmosphere (Chapter 9) (Ser-Giacomi et al., 2015c) we gain insights on the main pathways of transport in time-dependent geophysical flows.

Part I

THE FLOW NETWORK PARADIGM

CHAPTER 2

Flow network construction

2.1

Discretization and Lagrangian trajectories

Since fluid flow is a process occurring in continuous space, a discretization procedure involving a coarse-graining of space is needed to have access to the techniques of network theory. Advantages of the discrete point of view have already been shown in geophysical contexts (Froyland et al., 2007; Dellnitz et al., 2009; Santitissadeekorn et al., 2010; Froyland et al., 2012, 2014). Here we enumerate the steps needed to construct the discrete transport network starting from the continuous flow (Ser-Giacomi et al., 2015a).

Networks are composed of discrete building blocks: *nodes*. Being fluid flow a continuous system we need a discretized version of it to give a network representation. To do this we subdivide the fluid domain of interest in a large number N of boxes, $\{B_i, i = 1, ..., N\}$, so that network node j represents the fluid box B_j. Although it is not strictly necessary, we consider here the case in which boxes have the same area (in two-dimensional flows) or volume (for three dimensions). Then each box will contain exactly the same amount of fluid.

To complete the construction of our transport network, we need to establish the connections between nodes (i.e. boxes in the fluid domain). We establish a directional *link* between two nodes when an exchange of fluid occurred from one to the another during a given time interval. The *weight* of this link will be proportional to the amount of fluid transported. This quantity could be obtained from a Lagrangian point of view by following trajectories of ideal fluid particles and keeping record of their initial and final positions (i.e. starting and ending nodes) during the time interval considered.

More formally we integrate for a fixed time τ the equation of motion for each particle, from initial condition $\mathbf{x}_0$ at time t_0 until the final position $\mathbf{x}$ at $t_0 + \tau$, using a velocity field $\mathbf{v}(\mathbf{x}, t)$. This defines the *flow map* $\Phi_{t_0}^{\tau}$:

$$\mathbf{x}(t_0 + \tau) = \Phi_{t_0}^{\tau}(\mathbf{x}_0)\,, \tag{2.1}$$

which moves around single fluid particles. By considering the action of the flow map on all the points contained in a fluid region A we define the action of $\Phi_{t_0}^{\tau}$ on whole sets: $A(t_0 + \tau) = \Phi_{t_0}^{\tau}(A(t_0))$.

2.2

Adjacency matrix construction

Applying the flow map to the discrete boxes, we will have an estimation of the flow among each pair of nodes. More explicitly, given the collection of boxes $\{B_i, i = 1, ..., N\}$, we represent the transport between them by the discrete version of the Perron-Frobenious operator $\mathbf{P}(t_0, \tau)$, obtained within the Ulam approach, whose matrix elements are given by (Froyland and Dellnitz, 2003; Froyland, 2005; Froyland et al., 2007; Dellnitz et al., 2009; Froyland et al., 2010; Santitissadeekorn et al., 2010):

$$\mathbf{P}(t_0, \tau)_{ij} = \frac{m\left(B_i \cap \Phi_{t_0+\tau}^{-\tau}(B_j)\right)}{m(B_i)}, \tag{2.2}$$

$m(A)$ is a measure assigned to the set A. In our case it is the amount of fluid it contains, i.e. simply its area or volume. Other measures referring for example to heat or salt content could be implemented for future applications. Eq. (2.2) states that the flow from box B_i to box B_j is the fraction of the contents of B_i which is mapped into B_j. We refer to the figure in Appendix A for a plot of the different sets involved. If a nonuniform distribution of some conserved tracer is initially released in the system such that $\{p_i(t_0), i = 1, ..., N\}$ is the amount of such tracer in each box $\{B_i\}$ at the initial instant t_0, the matrix $\mathbf{P}(t_0, \tau)$ gives the evolution of this distribution after a time τ as $p_j(t_0 + \tau) = \sum_{i=1}^{N} p_i(t_0)\mathbf{P}(t_0, \tau)_{ij}$. Writing the $\{p_i\}$ as row vectors: $p(t_0 + \tau) = p(t_0)\mathbf{P}(t_0, \tau)$. A probabilistic interpretation of Eq. (2.2) is that $\mathbf{P}(t_0, \tau)_{ij}$ is the probability for a particle to reach the box B_j, under the condition that it started from a uniformly random position within box B_i. The matrix $\mathbf{P}(t_0, \tau)$ is row-stochastic, i.e. it has non-negative elements and $\sum_{j=1}^{N} \mathbf{P}(t_0, \tau)_{ij} = 1$, but not exactly column stochastic. The quantity $\sum_{i=1}^{N} \mathbf{P}(t_0, \tau)_{ij}$ measures the ratio of fluid present in box B_j after a time τ with respect to its initial content at time t_0. This ratio will be unity, and the matrix doubly stochastic, if the flow $\mathbf{v}(\mathbf{x}, t)$ is incompressible.

As a standard way to evaluate numerically the matrix in Eq. (2.2) we apply the Lagrangian map to a large number of particles released uniformly inside each of the boxes $\{B_i, i = 1, ..., N\}$ (see Fig. 2.1). The initial number of particles N_i in each box, a proxy of the amount of fluid it contains, should be proportional to its measure $m(B_i)$ which, with our choice of equal area or volume, results in seeding the same number of particles in each box. The number of particles transported from box B_i to box B_j gives an estimation of the flow among these boxes, and a numerical approximation to Eq. (2.2) is then:

$$\mathbf{P}(t_0, \tau)_{ij} \approx \frac{\text{number of particles from box } i \text{ to box } j}{N_i}. \tag{2.3}$$

Because of the time-dependence of the velocity field, the results of the Lagrangian simulations will depend on both the initial time t_0 and the duration of the simulation τ. Once these parameters are fixed, we can build a network described by a transport matrix $\mathbf{P}(t_0, \tau)$ that characterizes the connections among each pair of nodes from initial time t_0 to final time $t_0 + \tau$. We interpret $\mathbf{P}(t_0, \tau)$ as the adjacency matrix of a weighted and directed network, so that $\mathbf{P}(t_0, \tau)_{ij}$ is the weight of the link from node i to node j.

The network constructed in this way characterizes the final locations of all fluid elements a time τ after their release at time t_0, but gives no information on particle locations at

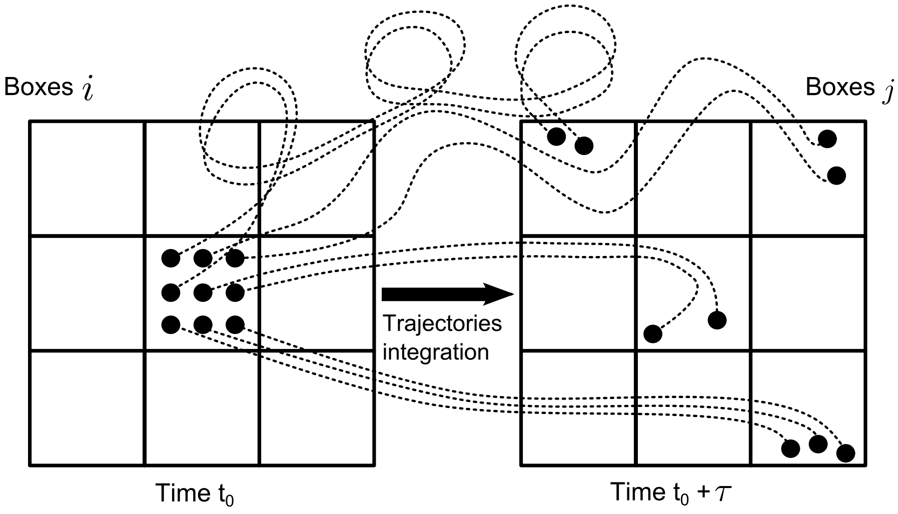

Figure 2.1: Transport matrix construction from tracer's advection, following Eq. (2.3). Picture from (Ser-Giacomi et al., 2015a).

intermediate times. Also, since each of the matrices $\mathbf{P}(t_0+k\tau,\tau)$, for $k=0,1,...,n-1$, is a stochastic matrix, one can consider the discrete-time Markov chain in which an initial vector giving occupation probabilities $p(t_0)=(p_1(t_0),...,p_N(t_0))$ for the different boxes is evolved in time as $p(t_n)=p(t_0)\mathbf{P}(t_0,\tau)\mathbf{P}(t_1,\tau)...\mathbf{P}(t_{n-1},\tau)$, where $t_k=t_0+k\tau$. This time evolution will not be exactly equal to the true evolution $p(t_n)=p(t_0)\mathbf{P}(t_0,n\tau)$, but a Markovian approximation to it in which the memory of the particle positions is lost after a time τ. The Markovian approximation may be reasonable in some circumstances and in fact it has been successfully used in geophysical flow problems (Dellnitz et al., 2009; Froyland et al., 2012, 2014). Note that here we are (and through Parts I and II of this thesis) not assuming any Markovian hypothesis and we work with the full matrix $\mathbf{P}(t_0,\tau)$ covering our time interval of interest and describing only the initial and final states of the transport process.

Despite not using any Markov assumption, replacing the continuous flow system by a discrete network introduces discretization errors. Even if the integration is done accurately, the initial and final locations of the transported particles are only specified up to a precision Δ, given by the linear side of the boxes. This implies that our network approach does not display explicitly fluid structures smaller than the box length-scale Δ.

The simulation provides the initial and final positions for each particle, allowing us to compute the transport matrix $\mathbf{P}(t_0,\tau)$ from Eq. (2.3). A directed link is established from node i to node j if and only if $\mathbf{P}(t_0,\tau)_{ij}$ is non-vanishing. In that case its value gives the weight of such a link. Due to numerical limitations, some trajectories end up prematurely by "beaching" onto land areas outside of the partition $\{B_i\}$. Then, the denominator N_i in Eq. (2.3) is taken as the number of particles still in the sea at the end of the integration time τ. Since the beaching effect is small, affecting less than 5% of all particles in the longest simulations presented here (and only for near-shore boxes), we still assume in the following that the convenient equal-area condition remains approximately valid.

Note that the Lagrangian integration is done under the full resolution of the velocity field (see next Section). This means that particle trajectories contain the small-scale features produced by the model during time τ. While such details are not explicitly present in the network description $\mathbf{P}(t_0, \tau)$ after coarse-graining the initial and final positions to the box size Δ, their effects have been incorporated in a statistical way.

2.3

Mediterranean Sea setup

We now apply the previous general procedures to build and analyze the flow network associated to a realistic surface flow in the Mediterranean sea. The input velocity field originates from the Mediterranean Forecasting System Model (physics reanalysis component). It is a hydrodynamic model based on NEMO-OPA (Nucleus for European Modelling of the Ocean-PArallelisé, version 3.2 (Madec et al., 2008)) with a variational data assimilation scheme. It is a primitive equations model in spherical coordinates, implemented in the Mediterranean at $\frac{1}{16}$ deg horizontal resolution and 72 unevenly spaced vertical levels (Oddo et al., 2009). We use here the "Physics reanalysis" component for years $2002 - 2011$ downloaded from MyOcean website. The model covers entirely the Mediterranean Basin and extends into the Atlantic in order to better resolve the exchanges with the Atlantic Ocean at the Strait of Gibraltar. It is nested, in the Atlantic, within the monthly mean climatological fields computed from ten years of daily output of the $\frac{1}{4}$ deg global model (Drévillon et al., 2008). Details on the nesting technique and major impacts on the model results can be found elsewhere (Oddo et al., 2009). The model uses vertical partial cells to fit the bottom depth shape. It is forced by momentum, water and heat fluxes interactively computed by bulk formulae using the 6-h, 0.25 deg horizontal-resolution operational analysis and forecast fields from the European Centre for Medium-Range Weather Forecasts. Air-sea processes predict surface temperature (Tonani et al., 2008), while the water balance is computed as Evaporation minus Precipitation and Runoff. The evaporation is derived from the latent heat flux; the precipitation and the runoff are provided by monthly mean datasets. The Dardanelles inflow is parameterized as a river using climatological net inflow rates (Kourafalou and Barbopoulos, 2003). The data assimilation system is the OCEANVAR scheme (Dobricic and Pinardi, 2008). The background error correlation matrices, estimated from the temporal variability of parameters in a historical model simulation, vary seasonally in the sub-regions of the Mediterranean Sea characterized by different physical characteristics (Dobricic et al., 2007). The Mean Dynamic Topography is used for the assimilation of Sea Level Anomaly (SLA) (Dobricic, 2005). The assimilated data include: along track SLA, satellite Sea Surface Temperature (SST), in-situ temperature profiles by eXpandable Bathy Thermograph, in-situ temperature and salinity profiles by Argo floats, in-situ temperature and salinity profiles from Conductivity-Temperature-Depth casts. Objective Analyses of SST data are used for the correction of surface heat fluxes with the relaxation constant of 40 $Wm^{-2}K^{-1}$

We used daily horizontal velocity fields generated by the model in the whole Mediterranean basin during 10 years of simulation $(2002 - 2011)$ selecting only one layer at a nominal depth of 7.9 m. This layer extends in fact between 4.58 and 11.55 m depth, so that it has a vertical extension of 6.97 m. For the integration time scales used here (val-

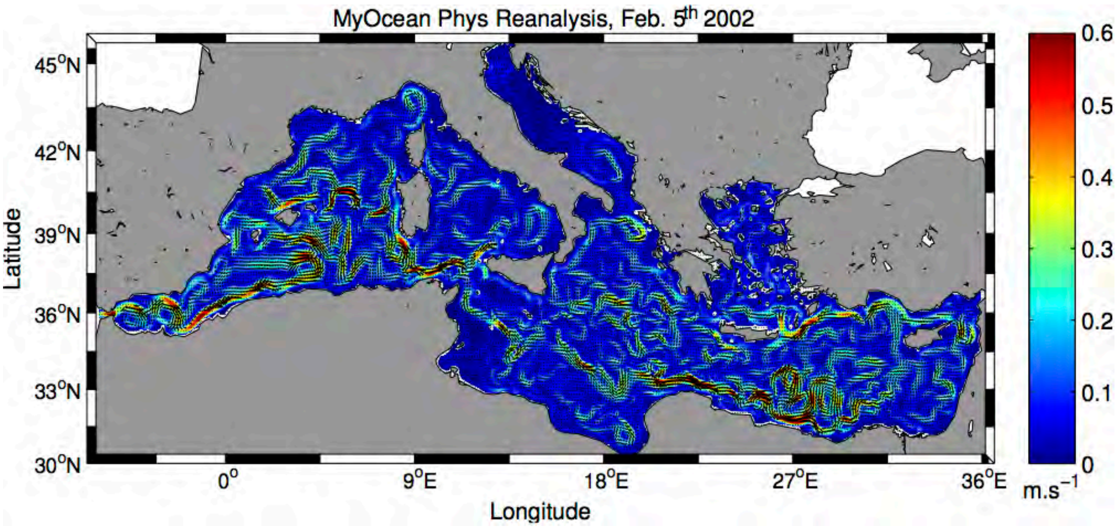

Figure 2.2: Snapshot of the surface (7.9 m) velocity field for February 5th 2002. Background colors represent the modulus (ms^{-1}) of the instantaneous velocity field.

ues of τ always below three months) we can reasonably neglect motion to other layers and consider only horizontal dynamics (d'Ovidio et al., 2004; Rossi et al., 2014). An exemplary snapshot of the surface (7.9 m) velocity field generated by this configuration and used to integrate Lagrangian particle trajectories is displayed in Fig. 2.2. Note the realistic representation of both large- and small-scale oceanographic features.

To switch from continuous space to discrete nodes we partition the above-described horizontal near-surface Mediterranean layer into 3270 two-dimensional square boxes. We imposed the equal-area constraint defining the cells in a sinusoidal projection given by coordinates x and y related to the standard longitude φ and latitude ϕ by

$$x = \varphi \cos\phi \qquad ; \qquad y = \phi \,. \tag{2.4}$$

In these x, y coordinates, boxes are squares of side 0.25 degrees, or $\Delta = 27.78\ km$ (see Fig. 2.3). The area Δ^2 of each box is 771.9 km^2. The "amount of water in a box B_i" is then related to its area Δ^2 through a simple multiplication by the layer thickness (6.97 m), returning a value of $5.38 \times 10^9\ m^3$ per box. The resolution of the model-generated velocity field is much finer than the discretization we use for network construction. In this sense the dynamics represented in the flow network is a coarse-graining of the simulated Mediterranean flow, keeping the effect of the small scales only in a statistical sense. The most energetic features of the Mediterranean flow are mesoscale structures (Millot and Taupier-Letage, 2005) ranging from 10 km to a few hundred km. With the value of Δ we use, our network description displays most of the mesoscale range, and neglects submesoscales, which anyway are only marginally resolved by the NEMO implementation.

To characterize the transport phenomena, $N_i = 500$ ideal fluid particles were released in each box B_i. We simulated the motion of these $3270 \times 500 = 1.635 \times 10^6$ particles by integrating the trajectories in the velocity field using a fourth-order Runge-Kutta algorithm. The velocity at any arbitrary point in the sea is computed with a bilinear interpolation from the input data. We used a time step of 1 day (the same resolution as

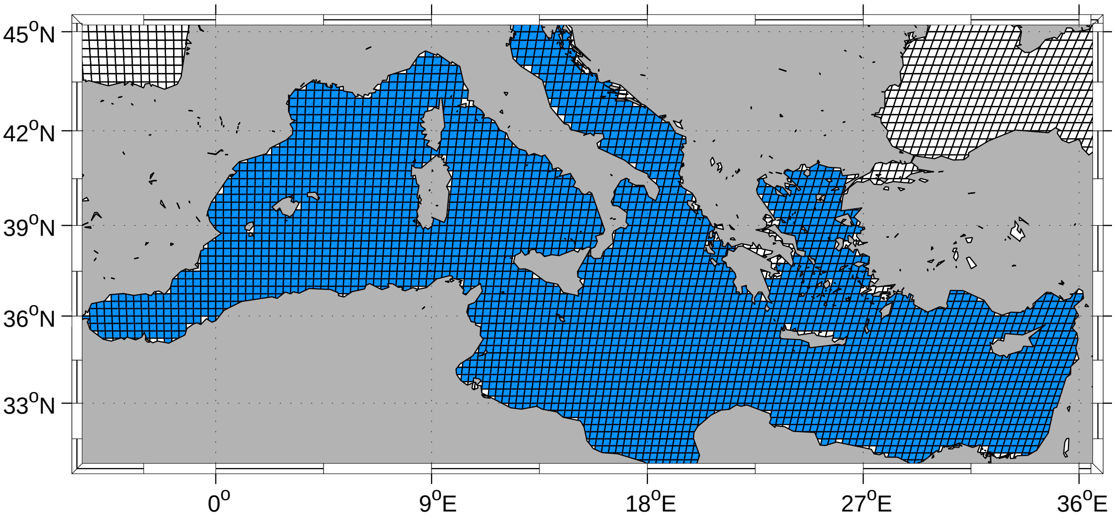

Figure 2.3: Discretization of the Mediterranean sea (blue region) into $N = 3270$ equal-area boxes $\{B_i, i = 1, ..., N\}$.

the data). We also tested shorter intervals using a cubic interpolation but no significant improvement was found. The two key-parameters of the simulations are the starting time t_0 and the tracking time τ.

CHAPTER 3

Local network properties

We now interpret the transport matrices $\mathbf{P}(t_0, \tau)$, for several values of t_0 and τ, as the adjacency matrices of directed and weighted flow networks. We can calculate for them all the standard quantities characterizing the topology of networks, such as degree, clustering, betweenness, etc. (Newman, 2009). But following the aim stated in the Introduction, we will concentrate here in network quantities that can give insight in (horizontal) dispersion and mixing processes, and in the identification of coherent regions.

3.1

Dispersion and mixing

Important properties of geophysical flows depend on their dispersion characteristics, i.e. how far away can the fluid be transported during some time, and how diverse are the target regions. Mixing of fluid with different characteristics, another process of great geophysical relevance, will occur at a particular place if fluid from different origins arrives there at a particular time.

In dynamical systems approaches to flow processes, a standard way to quantify dispersion is by means of the finite-time Lyapunov exponent (FTLE). It is defined as (Shadden et al., 2005)

$$\lambda(\mathbf{x}_0, t_0, \tau) = \frac{1}{|\tau|} \log \sqrt{\sigma(\mathbf{x}_0, t_0, \tau)}\,, \tag{3.1}$$

where $\sigma(\mathbf{x}_0, t_0, \tau)$ is the maximum eigenvalue of the Cauchy-Green strain tensor:

$$C(\mathbf{x}_0, t_0, \tau) = \left(\nabla\Phi_{t_0}^{\tau}(\mathbf{x}_0)\right)^T \nabla\Phi_{t_0}^{\tau}(\mathbf{x}_0)\,, \tag{3.2}$$

constructed from the Jacobian matrix $\nabla\Phi_{t_0}^{\tau}(\mathbf{x}_0)$ of the flow map. M^T means the transpose of the matrix M. For $\tau > 0$ this is the forward FTLE. By running the flow map backwards in time ($\tau < 0$) we get the backwards FTLE field, which quantifies the strength of mixing into a particular location. The interpretation of (3.1) is that an initial circle of infinitesimal diameter δ located at $\mathbf{x}_0$ at t_0 will become an ellipse of major axis $e^{\tau\lambda(\mathbf{x}_0, t_0, \tau)}\delta$ after being advected by the flow during a time τ. The minor axis will be a decreasing function of τ, contracting at an exponential rate related to a negative Lyapunov exponent that can be computed from the second eigenvalue of $C(\mathbf{x}_0, t_0, \tau)$.

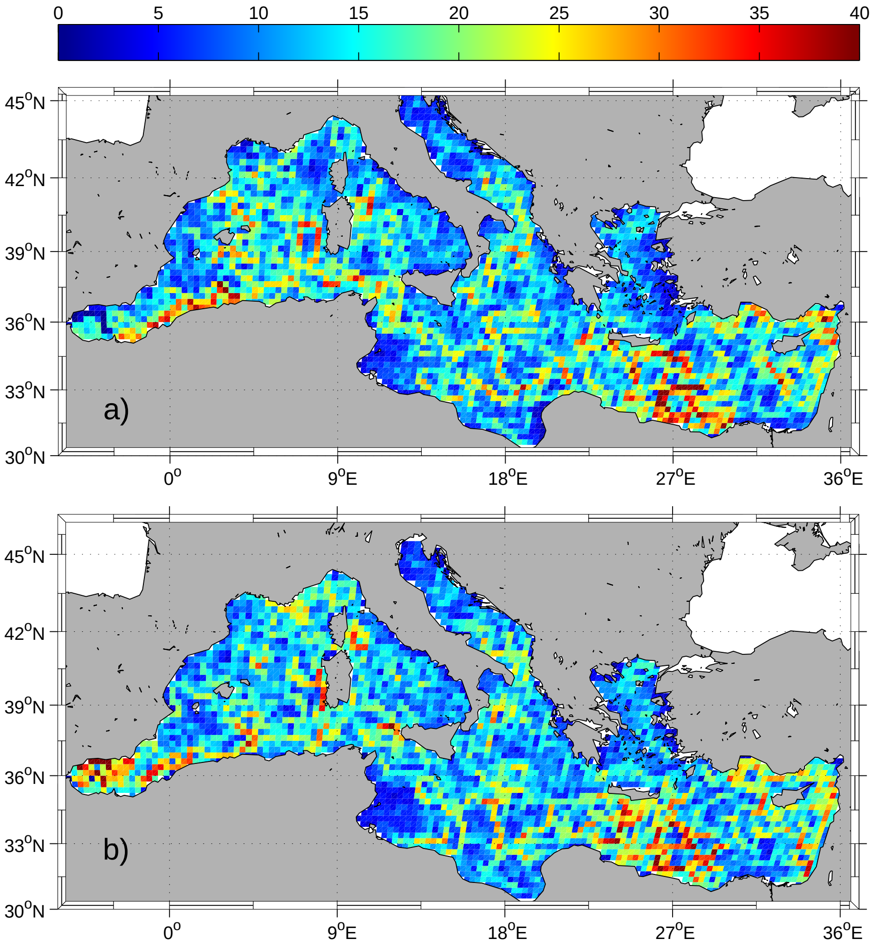

Figure 3.1: Degree of the nodes in the flow network defined by $\mathbf{P}(t_0, \tau)$, for t_0=July 1st 2011 and $\tau = 15$ days. a) The in-degree $K_I(i)$. b) The out-degree $K_O(i)$.

An obvious quantity in the network interpretation suitable to be related to dispersion and mixing is the *degree* of a node. Since our network is directed, we should distinguish between the in-degree $K_I(i)$, i.e. the number of links pointing to a particular node i, and the out-degree $K_O(i)$, the number of links pointing out of it. Figure 3.1 displays these quantities at the geographical locations defined by the nodes of the Mediterranean network for particular values of t_0 and τ. High values of the degrees appear associated to the strong and unstable currents in the southern part of the basin (Millot and Taupier-Letage, 2005). Low degree values are observed in regions where the circulation is rather slow, such as the Tunisian continental shelf and the semi-enclosed seas (e.g. Adriatic and Aegean). Generally, the values of the in- and out degree- tend to increase with τ. With respect to the dependence on t_0, degree values tend to be slightly higher in winter than in summer.

A first problem in relating these network properties to the actual physics of dispersion and mixing is that their values are dependent on the spatial scales chosen for discretization (there is also a dependency on the numbers N_i of particles used to compute the transport matrices, but it disappears for large N_i). This problem is easy to solve by recalling that every box has an associated area. Dealing first with the out-degree case for definiteness, $K_O(i)$ is proportional to the total area of all nodes that received some contents from the initial node i. This quantity has a well-defined meaning that can be related to the continuous flow dynamics with only a minor dependence on the discretization procedure. Since here all boxes have the same area Δ^2, the area corresponding to the out-degree of node i is $K_O(i)\Delta^2$. We can use generic ideas of chaotic dynamics to obtain heuristically a more precise relationship between two quantifiers of dispersion: the degree and the Lyapunov exponent. In regions dominated by hyperbolic structures, each of the fluid boxes will be stretched into a long and thin filament after a sufficiently long time τ (see Appendix A). If we want to compute the number of boxes reached by it, it is enough to consider its length, since the width quickly becomes smaller than the box size Δ. Let us consider an initial line of length $L(t_0) \approx \Delta$ inside the initial box B_i. A small segment of it, of length $dl(t_0)$ at position $\mathbf{x}_0 \in B_i$ will become elongated by a factor given by the local FTLE: $dl(t_0+\tau) = dl(t_0)e^{\tau\lambda(\mathbf{x}_0,t_0,\tau)}$. Integrating over the initial positions along the line we get an estimation of the final length $L(t_0+\tau)$ of the filament. A better estimation $\bar{L}(t_0+\tau)$ of this length can be done by averaging over positions transverse to the line, to take into account different locations of the initial line in the box:

$$\bar{L}(t_0+\tau) \approx \frac{1}{\Delta}\int_{B_i} d\mathbf{x}_0 e^{\tau\lambda(\mathbf{x}_0,t_0,\tau)}\,, \tag{3.3}$$

where the longitudinal and transverse integrations have been combined into the integration of $\mathbf{x}_0$ over the area B_i. The area of the boxes covered by the filament is $\mathcal{A}(t_0+\tau) \approx \bar{L}(t_0+\tau)\Delta$ so that the out-degree of the initial box will be

$$\begin{aligned} K_O(i) = \frac{\mathcal{A}(t_0+\tau)}{\Delta^2} &\approx \quad , \\ \frac{1}{\Delta^2}\int_{B_i} d\mathbf{x}_0 e^{\tau\lambda(\mathbf{x}_0,t_0,\tau)} &\equiv \left\langle e^{\tau\lambda(\mathbf{x}_0,t_0,\tau)} \right\rangle_{B_i}\,. \end{aligned} \tag{3.4}$$

Thus, we have a useful relationship between a natural quantity in the network description of fluid flows and a standard characterization of dispersion in the dynamical systems approach to such flows: the degree of a node associated to a box is the average or coarse-graining of the stretching factor $e^{\tau\lambda}$ in that box.

We can check the validity of the above heuristic arguments by comparing directly the values of $K_O(i)$ obtained from our flow network and the right-hand-side of (3.4). Figure 3.2 shows an example of FTLE field obtained at time t_0 = July 1st 2011, and $\tau = 15$ days. Figure 3.3 shows the clear correlation between the two quantities. Three values of τ are plotted to appreciate the general validity of the relationship. We attribute the deviations with respect to the exact identity to the fact that the filament-type arguments are only valid for sufficiently large τ and in regions dominated by strain. Also, our arguments neglect the presence of filament foldings that sometimes would occupy the same box, and of associated saturation effects. In addition quantization effects arising from the discrete nature of K_O are visible at small degree values.

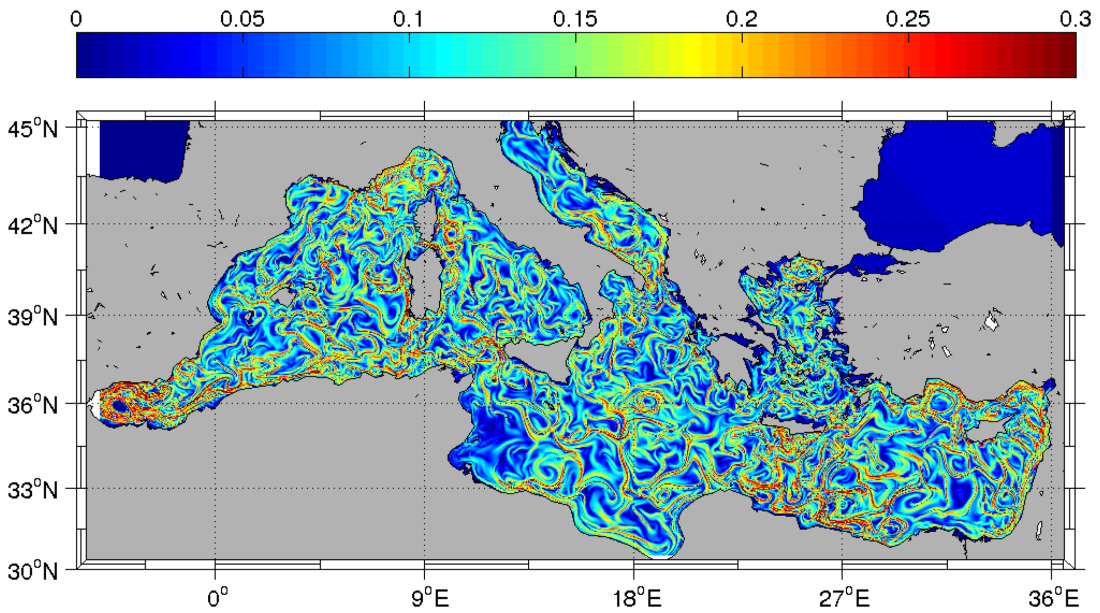

Figure 3.2: An example of forward FTLE field $\lambda(\mathbf{x}_0, t_0, \tau)$ at t_0 =July 1st 2011, and $\tau = 15$ days. Color bar in day^{-1}

3.2

Network entropies

Expression (3.4) suggests defining:

$$H_i^0(t_0,\tau) \equiv \frac{1}{\tau}\log K_O(i)\,, \tag{3.5}$$

so that:

$$\left\langle e^{\tau\lambda(\mathbf{x}_0,t_0,\tau)}\right\rangle_{B_i} = e^{\tau H_i^0(t_0,\tau)}\,. \tag{3.6}$$

From the convexity of the exponential function, we have $H_i^0(t_0,\tau) \geq \langle\lambda(\mathbf{x}_0,t_0,\tau)\rangle_{B_i}$. The previous expressions are reminiscent of the properties of the topological entropy of a dynamical system, as giving the exponential growth in time of the length of a material line (Tél and Gruiz, 2006). Pushing forward the analogy, we can define a sequence of Rényi-like entropies (Rényi, 1970) associated to a particular node i:

$$H_i^q(t_0,\tau) \equiv \frac{1}{(1-q)|\tau|}\log\sum_{j=1}^{N}\left(\mathbf{P}(t_0,\tau)_{ij}\right)^q\,, \tag{3.7}$$

which we call *network entropies*. Due to their dependence on the finite-size of the partition, they are related to the ϵ-entropies discussed by (Boffetta et al., 2002). Note however that here the transport matrix involves only two states of the trajectories, separated by an interval of time τ which remains finite, and the dependence on the initial location, box B_i, is kept. The entropies H_i^0 and H_i^1 should be understood as defined by the limits $q \to 0$ and $q \to 1$, respectively. All the network entropies measure the *diversity* in the amounts of fluid received by the nodes connected to a given box, but

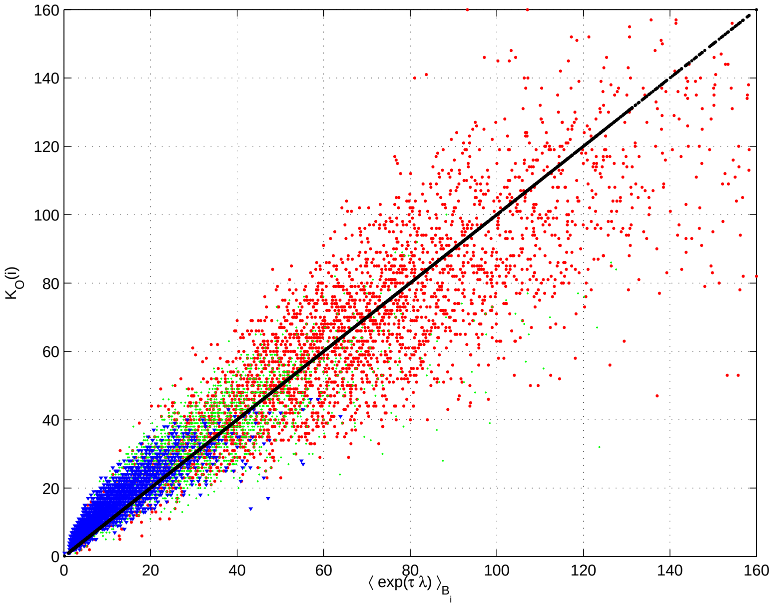

Figure 3.3: Values of the out-degree $K_O(i)$ of each node i vs the average value of the stretching factor $e^{\tau\lambda}$ in that node. t_0 =July 1st 2011. Blue symbols are from $\tau = 15$ days, green from $\tau = 30$ days and red from $\tau = 60$ days. Black line is the main diagonal.

weighting them in different ways: In H_i^0 all nodes are counted equally independently of the amount of water they receive, so that it informs only about the degree as seen in Eq. (3.5); for increasing values of q nodes receiving more water are weighted with increasing strength. Although the network entropies have been introduced here in the particular context of flow networks, we note that they can be defined for any weighted network, giving generalizations of the degree to quantify the unevenness of the weight distribution towards the nodes connected to a given one.

Applying l'Hôpital's rule to the definition of the network entropy of order $q = 1$ one gets:

$$H_i^1(t_0, \tau) = -\frac{1}{\tau} \sum_{j=1}^{N} \mathbf{P}(t_0, \tau)_{ij} \log \mathbf{P}(t_0, \tau)_{ij} \, . \tag{3.8}$$

It gives the amount of information (per unit of time) gained by observing the position of a particle at time $t_0 + \tau$, knowing that it was initially (time t_0) somewhere in box B_i. This quantity is precisely the discrete finite-time entropy studied by (Froyland and Padberg-Gehle, 2012). Figure 3.4 shows its spatial distribution in the Mediterranean sea for particular values of t_0 and τ.

The standard Pesin-like results relating the metric or Kolmogorov-Sinai entropy to the sum of positive Lyapunov exponents (Boffetta et al., 2002; Castiglione et al., 2010; Cencini et al., 2010) suggest that, at least for large τ, the entropy H_i^1 would give a good approximation to the values of the FTLE field averaged over each box B_i: $\lambda_i(t_0, \tau) \equiv \langle \lambda(\mathbf{x}_0, t_0, \tau) \rangle_{B_i} \approx H_i^1(t_0, \tau)$. Appendix A gives calculations supporting this claim in an heuristic way. Figure 3.5 shows the geographical distribution of $\lambda_i(t_0, \tau)$ and Fig. 3.6 compares both quantities for several values of τ. The entropies tend to be slightly larger

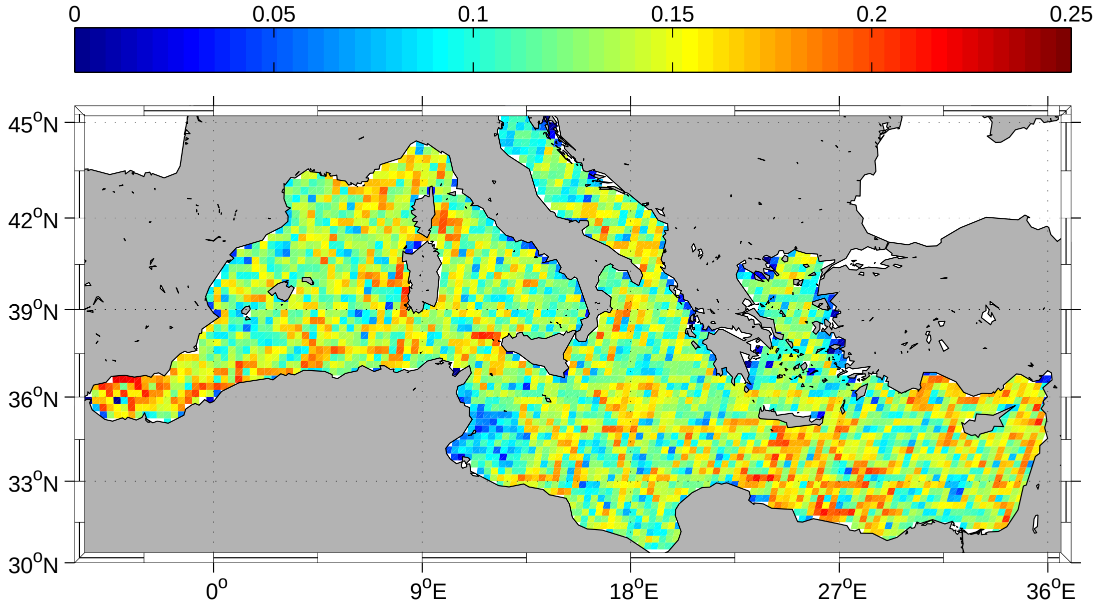

Figure 3.4: The network entropy $H_i^1(t_0, \tau)$, for t_0 =July 1st 2011, and $\tau = 15$ days. Color bar in day^{-1}.

than the Lyapunov exponents for $\tau = 15$ days, but both quantities approach each other and become well correlated for larger τ.

For definiteness we have been discussing quantities related to the forward time evolution: out-degree, forward Lyapunov exponents, etc. The network entropies can also be defined for the backward time evolution. Construction of the backwards-dynamics network can be achieved by redoing the launching of particles and running the Lagrangian integration for negative time, or much simpler, by recognizing (Froyland and Padberg-Gehle, 2012) that the backward evolution is given by the matrix

$$\mathbf{P}(t_0 + \tau, -\tau)_{ij} = \frac{\mathbf{P}(t_0, \tau)_{ji}}{\sum_{k=1}^{N} \mathbf{P}(t_0, \tau)_{ki}} . \qquad (3.9)$$

The network entropies in Eq. (3.7) can now be directly computed for the backward flow network defined by $\mathbf{P}(t_0 + \tau, -\tau)$, and they will be related to backwards Lyapunov fields, which give a measure of mixing of fluid coming from different origins. As an example we show in Fig. 3.7 the relationship between the backwards entropy $H_i^1(t_0 + \tau, -\tau)$ and the coarse-grained backwards Lyapunov exponent $\lambda_i(t_0 + \tau, -\tau)$. Again both quantities are similar for sufficiently large τ and the same qualitative features as in Fig. 3.6 are observed.

Summarizing this Chapter, we have defined a family of entropy-like quantities completely in terms of the transport matrix characterization of the flow network. At least two of them, H_i^0 and H_i^1, are related to standard dispersion and mixing quantifiers in the description of fluid flows. The higher order entropies H_i^q are related to the generalized Lyapunov exponents (Boffetta et al., 2002; Cencini et al., 2010) characterizing successive moments of the Lyapunov field, as discussed in Appendix A. We do not claim that these relationships are exact for finite values of τ and Δ. Instead, we find numerical deviations from them (Figs. 3.3, 3.6 and 3.7) which decrease for increasing τ. We expect

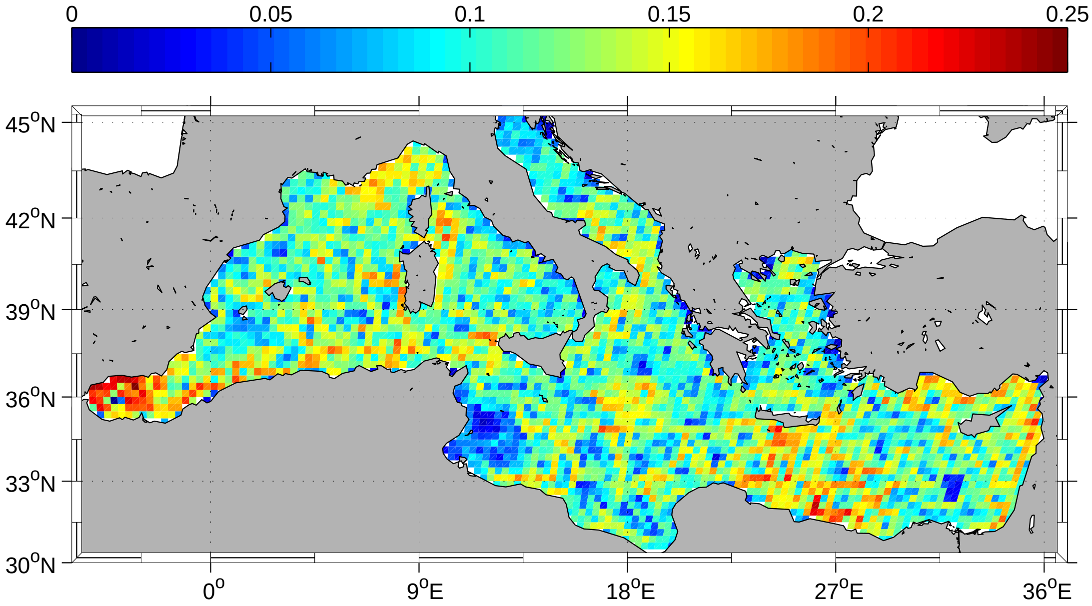

Figure 3.5: Coarse-graining of the Lyapunov field in Fig. 3.2 into the discretization boxes: $\lambda_i(t_0, \tau) \equiv \langle \lambda(\mathbf{x}_0, t_0, \tau) \rangle_{B_i}$. t_0 =July 1st 2011, and $\tau = 15$. Color bar in day^{-1}.

the same to happen when decreasing Δ. Recent works also showed similar results in theoretical two-dimensional flows (Lindner, 2015) showing a striking matching when using a sufficiently small Δ. The important point is that, once the network matrix $\mathbf{P}(t_0, \tau)$ has been constructed, the entropies in Eq. (3.7) provide a computationally very cheap way to assess quantities of geophysical interest such as local dispersion, stretching and mixing. In fact the simplest network quantifiers such as the in- and out-degrees are already suitable for that, being related to H_i^0. The qualitative information displayed in figures such as 3.1b and 3.4 or 3.5 is very similar. Also, even if we should have $H_i^0 \geq H_i^1 \approx \lambda_i$, in our examples the numerical values of H_i^0 are only slightly larger than those of H_i^1. We have to mention that we have been working under the hypothesis of boxes $\{B_i\}$ of equal areas. Expression (3.7) needs indeed corrections when dealing with differing area boxes (see Appendix B).

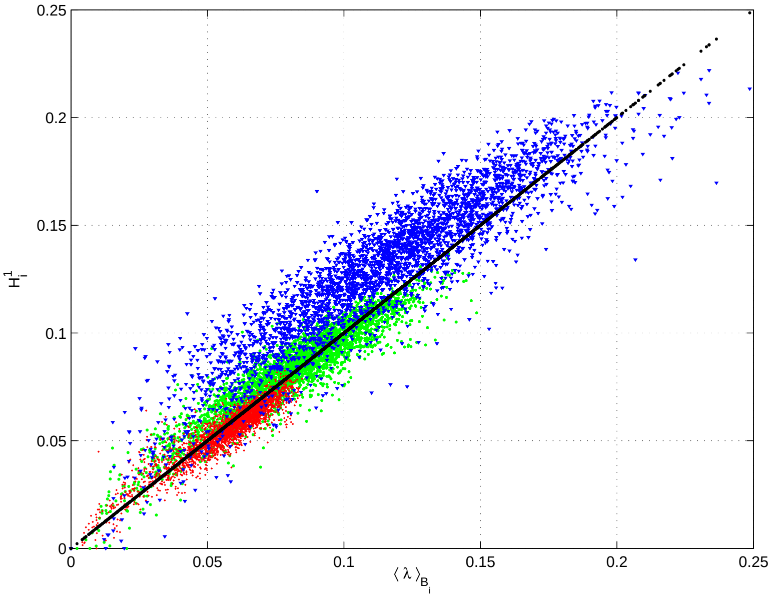

Figure 3.6: Values of the network entropy $H_i^1(t_0, \tau)$ of each node i vs the average value of the Lyapunov exponent in that node, $\lambda_i(t_0, \tau)$. t_0 =July 1st 2011. Blue symbols are from $\tau = 15$ days, green from $\tau = 30$ days and red from $\tau = 60$ days. Black line is the main diagonal.

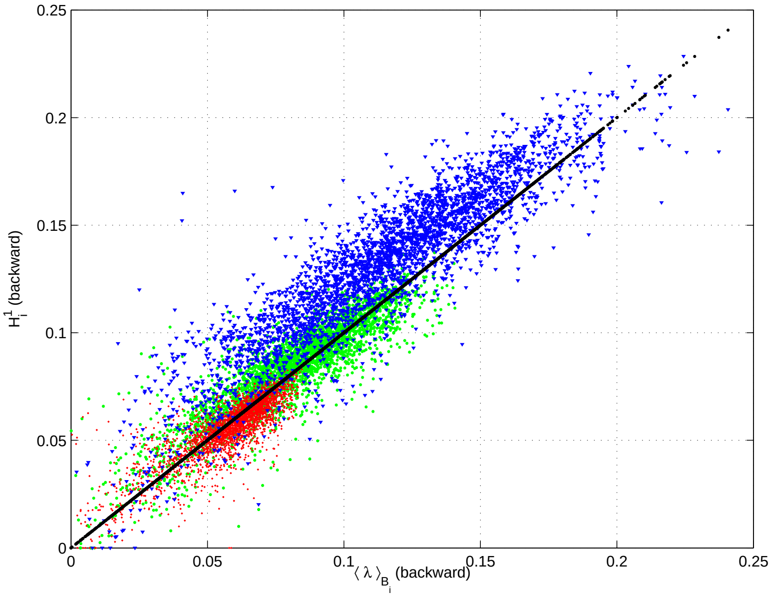

Figure 3.7: Values of the network entropy H_i^1 at each node i, computed from the backwards-dynamics network given by $\mathbf{P}(t_0+\tau, -\tau)$ (Eq. (3.9)), vs the average value of the backwards Lyapunov exponent in that node, $\lambda_i(t_0+\tau, -\tau)$. t_0 =July 1st 2011. Blue symbols are from $\tau = 15$ days, green from $\tau = 30$ days and red from $\tau = 60$ days. Black line is the main diagonal.

CHAPTER 4

Community detection

4.1

Coherent regions as network communities

Most work in the dynamical systems approach to fluid transport aims at identifying "barriers to transport" locating the borders of regions that do not exchange much fluid among them. The set-oriented approach focusses on the regions themselves rather than on the borders. Almost-invariant fluid regions have been defined as regions of the fluid domain remaining relatively isolated (according to a suitably defined metrics) from the rest of the fluid (Froyland and Dellnitz, 2003; Froyland, 2005). In generic time-dependent flows these regions will not be fixed in space but they will be transported by the mean flow, and the concept of coherent pairs, relating initial and final set positions has been developed (Froyland et al., 2010; Santitissadeekorn et al., 2010; Froyland et al., 2012). Formulating this problem in the context of network theory would require building on techniques for bipartite graphs. In our present case study, the global flow in the Mediterranean sea, land masses play an important role in restricting the flow, so that coherent regions that remain fixed with respect to the coasts are the most relevant ones for many applications. In particular, when considering environmental conservation strategies and marine reserves (Nilsson-Jacobi et al., 2012; Thomas et al., 2014; Rossi et al., 2014), one looks for the connectivity among marine zones, or provinces (Rossi et al., 2014), occupying localized regions of the sea. Thus we focus here on finding a partition of the sea into *self-coherent*, or *almost-invariant* regions, associated to relatively stable circulation patterns, from the point of view of network theory. We want these regions to be well-mixed internally, and with little interchange with the exterior. In the language of networks this translates to partitioning the network into subgraphs with high internal connectivity, and small connectivity among them. This is the standard problem of community detection in networks (Newman, 2009; Danon et al., 2005; Lancichinetti and Fortunato, 2009; Fortunato, 2010; Aldecoa and Marín, 2013), for which many different and powerful techniques are available. In fact, most of the approaches used so far to partition fluid motion into almost-invariant sets (Froyland and Dellnitz, 2003; Froyland, 2005; Speetjens et al., 2013) employ classical spectral techniques for graph partition (Newman, 2009), which use the eigenvectors or singular vectors of the transport matrix (or other matrices derived from it). We note that the methodologies in (Froyland and Dellnitz, 2003; Froyland, 2005) find almost-invariant sets in the sense that loss and gain of fluid is minimized. But the condition of strong internal mixing, which we consider important in geophysical applications, is not imposed.

Here we address the community detection problem with a state-of-art network-theory approach, the *Infomap* algorithm (Rosvall and Bergstrom, 2008). The method is based on the probability flow of random walks in the network moving with transition probabilities given by the adjacency matrix $\mathbf{P}(t_0, \tau)$, and on exploiting the properties of information compression in the description of that probability flow. *Infomap* finds the partition of the network minimizing the average size of the codeword needed to describe inter- and intracommunity transitions (a succinct description of the method is provided in Section 2.3). We believe this methodology is specially suited to partition flow networks for the following reasons: First, it takes into account the "direction" and "weight" of each link, important characteristics defining our flow network. The standard spectral methods and most modularity-optimization algorithms take as input a symmetrized version of the network. Second, *Infomap* does not require to fix *a priory* the number of communities forming the domain partition. Third, *Infomap* does not impose similar sizes to the communities so that it does not suffer from the "resolution limit" (Fortunato and Barthélemy, 2007) restricting the minimum community size detectable by most algorithms, including spectral methods. This is important in geophysical flow networks since ocean structures of different sizes coexist in the sea, some of them arising from geographical accidents, bathymetry, etc.

The method has also some limitations. One of them is the "field of view limit" (Schaub et al., 2012) due to the use of a single-step transition matrix $\mathbf{P}(t_0, \tau)$. In general this imposes that the detected communities are only those with intense intracommunity connections (clique-like). For our application this feature may become convenient since *Infomap* will identify as communities only regions well mixed internally by the flow.

Since *Infomap* consider random walkers exploring the network with the transition probabilities in the matrix $\mathbf{P}(t_0, \tau)$, one is tempted to confuse these walkers with the Lagrangian particles advected by the flow. But this is not correct. $\mathbf{P}(t_0, \tau)$ contains relationships between initial and final positions of particles after a time τ, but does not describe in detail the trajectories at intermediate times. In addition it can not be used beyond that time since in time-dependent velocity fields flow connectivity will change with the initial time t_0, defining the dynamic network. *Infomap* unveils the graph structures present in the single matrix $\mathbf{P}(t_0, \tau)$ by releasing random walkers that evolve in a virtual time not directly related to the physical time.

Hydrodynamical provinces delimited by *Infomap* in the Mediterranean surface flow will be further studied in Chapter 8 (Rossi et al., 2014), where we will discuss also their implications for the design of marine reserves. Here we concentrate in the technical aspects and compare with alternative methods.

4.2

Quality parameters

A standard way to asses the quality of a network partition is by computing a modularity parameter (Newman and Girvan, 2004; Newman, 2009). But this involves comparison with a random null model that in the case of flow networks has no obvious meaning. Then we prefer to use alternative quantifiers with a direct interpretation in terms of fluid connectivity. Here we define a *coherence ratio* and a *mixing parameter*.

If coherent regions A are understood as almost-invariant areas of fluid, this means that they are mapped by the flow nearly into themselves after a time τ:

$$\Phi_{t_0}^{\tau}(A) \approx A\,. \tag{4.1}$$

To measure how well this is achieved one can introduce the *coherence ratio* (Froyland and Dellnitz, 2003; Froyland, 2005):

$$\rho_{t_0}^{\tau}(A) = \frac{m(A \cap \Phi_{t_0+\tau}^{-\tau}(A))}{m(A)}\,, \tag{4.2}$$

where, as before, $m(C)$ is the area of set C, but it can be generalized to other measures. We have $\rho_{t_0}^{\tau}(A) \leq 1$ and values close to unity indicate that A is a truly almost-invariant set.

In our discrete set-up, we consider sets A made of our boxes $\{B_i, i = 1, ..., N\}$: $A = \cup_{i\in\mathcal{I}} B_i$, where $\mathcal{I}$ is the set of indices identifying the boxes B_i making A. The coherence ratio is now (Froyland and Dellnitz, 2003; Froyland, 2005)

$$\rho_{t_0}^{\tau}(A) = \frac{\sum_{i,j\in\mathcal{I}} m(B_i)\mathbf{P}(t_0,\tau)_{ij}}{\sum_{i\in\mathcal{I}} m(B_i)}\,. \tag{4.3}$$

For a partition of the fluid domain into p communities or provinces: $\mathcal{P} = \{A_1, ..., A_p\}$, a global quality figure of the partition is

$$\rho_{t_0}^{\tau}(\mathcal{P}) \equiv \frac{1}{p}\sum_{k=1}^{p} \rho_{t_0}^{\tau}(A_k)\,, \tag{4.4}$$

where again a good partition would be indicated by a value close to 1. When communities are of very different sizes it may be appropriate to weight the average in Eq. (4.4) with these sizes, but we keep the present definition to allow comparison with previous works.

Physically we can say that $\rho_{t_0}^{\tau}(\mathcal{P})$ represents the fraction of tracers that at time $t_0 + \tau$ are found in the same province where they were released at time t_0. The definition involves the initial and final positions, but gives no information on the particle trajectories in between. Note that coherence ratios measure fluid exchanges between provinces, but do not quantify how strong the internal mixing is.

The second quantifier we use is a *mixing parameter* devised to assess how strongly the flow mixes fluid inside communities. To define the mixing parameter $M_{t_0}^{\tau}(A)$ inside a set A we first define a transport matrix conditioned to represent just the transport occurring inside A (more precisely, transport by trajectories that start and end in A):

$$\mathbf{R}(t_0,\tau|A)_{ij} = \frac{\mathbf{P}(t_0,\tau)_{ij}}{\sum_{k\in\mathcal{I}} \mathbf{P}(t_0,\tau)_{ik}}\,, \quad i,j \in \mathcal{I}\,. \tag{4.5}$$

As before, $\mathcal{I}$ is the set of indices identifying the boxes B_i making A. The mixing parameter is a normalized version of the sum inside A of the entropies associated to the transition probabilities in $\mathbf{R}(t_0,\tau|A)$:

$$M_{t_0}^{\tau}(A) = \frac{-\sum_{i,j\in\mathcal{I}} \mathbf{R}(t_0,\tau|A)_{ij} \log \mathbf{R}(t_0,\tau|A)_{ij}}{Q_A \log Q_A}\,. \tag{4.6}$$

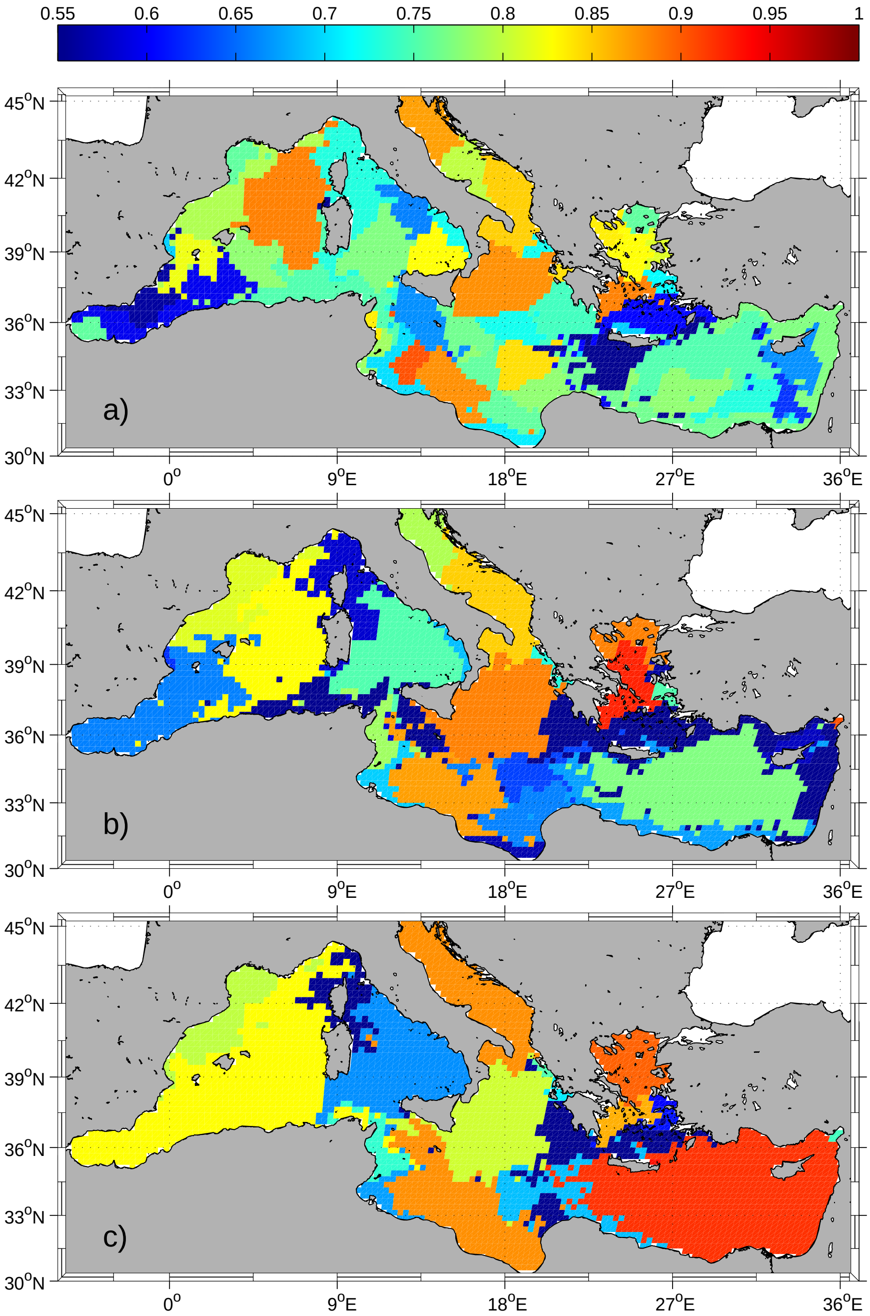

Figure 4.1: *Infomap* partition of flow networks in the Mediterranean sea, defined by $\mathbf{P}(t_0, \tau)$, into communities or provinces for increasing values of τ. Each province is colored by its coherence ratio value from Eq. (4.3), as given in the color bar. In all panels t_0 = July 1st 2011. a) $\tau = 30$ days; the number of communities is $p = 56$, the global coherence $\rho^{\tau}_{t_0}(\mathcal{P}) = 0.76$, and the global mixing $M^{\tau}_{t_0}(\mathcal{P}) = 0.47$. b) $\tau = 60$ days; $p = 33$, $\rho^{\tau}_{t_0}(\mathcal{P}) = 0.73$, $M^{\tau}_{t_0}(\mathcal{P}) = 0.54$. c) $\tau = 90$ days; $p = 22$, $\rho^{\tau}_{t_0}(\mathcal{P}) = 0.80$, $M^{\tau}_{t_0}(\mathcal{P}) = 0.59$.

Q_A is the number of boxes in A. The maximum value, $M_{t_0}^{\tau}(A) = 1$, is reached when fluid is dispersed from each box in A to all the others uniformly ($\mathbf{R}_{ij} = 1/Q_A, \forall i, j \in \mathcal{I}$). A global quantification of the internal mixing in a community partition $\mathcal{P} = \{A_1, ..., A_p\}$ is given by

$$M_{t_0}^{\tau}(\mathcal{P}) = \frac{\sum_{k=1}^{p} m(A_k) M_{t_0}^{\tau}(A_k)}{\sum_{k=1}^{p} m(A_k)} . \tag{4.7}$$

Here, we have weighted the different communities according to their size.

4.3

Communities in the Mediterranean surface flow

The outputs of the *Infomap* algorithm applied to the flow network defined by $\mathbf{P}(t_0, \tau)$, for increasing values of τ, are shown in Fig. 4.1. Each community A_k is colored with the value of its coherence ratio $\rho_{t_0}^{\tau}(A_k)$. We see that most coherence values are rather high. The global mixing parameter has only moderate values (see caption of Fig. 4.1), but it increases with τ. The main coast-constrained regions appear clearly outlined (the Tyrrhenian, the Adriatic, the Aegean, ...), but also other areas defined only by persistent circulation patterns (the three-gyre system in the Adriatic, the Balearic front, ...). We refer to (Rossi et al., 2014) and to Chapter 5 for a thorough interpretation of the hydrodynamic provinces in relation with surface circulation patterns and known eco-regionalization of the Mediterranean basin. Note that there is no obvious relationship between the size of a community and its coherence. Both large and small provinces may have indeed moderate (< 0.6) or high (> 0.8) coherence ratios. The detection of small communities confirms that *Infomap* is not affected by the "resolution limit" (Fortunato and Barthélemy, 2007).

Communities merge and in average become larger with increasing τ, so that their number decreases. Fig. 4.2a shows the growth of the mean area as a function of τ for the same case t_0 = July 1st 2011 shown in Fig. 4.1. The standard deviation of the area distribution is also displayed as error bars. It shows a significant dispersion in the area of the communities identified, especially for larger τ, revealing properly the multi-scale character of oceanic transport processes. For small τ, community areas seem mainly controlled by the time of integration (there is no sufficient time for the flow to manifest highly inhomogeneous dispersion) but only marginally determined by the intrinsic properties of the flow. As commented above, detecting communities of widely different sizes is a great capability of Infomap, whereas other methodologies constrain the communities to be of similar sizes. The inset Fig. 4.2b shows how the number of communities decreases when τ increases.

4.4

Average community descriptions

Because of the turbulent nature of oceanic motions, the community decomposition changes with t_0. Some communities (even of small size) are repeatedly observed while

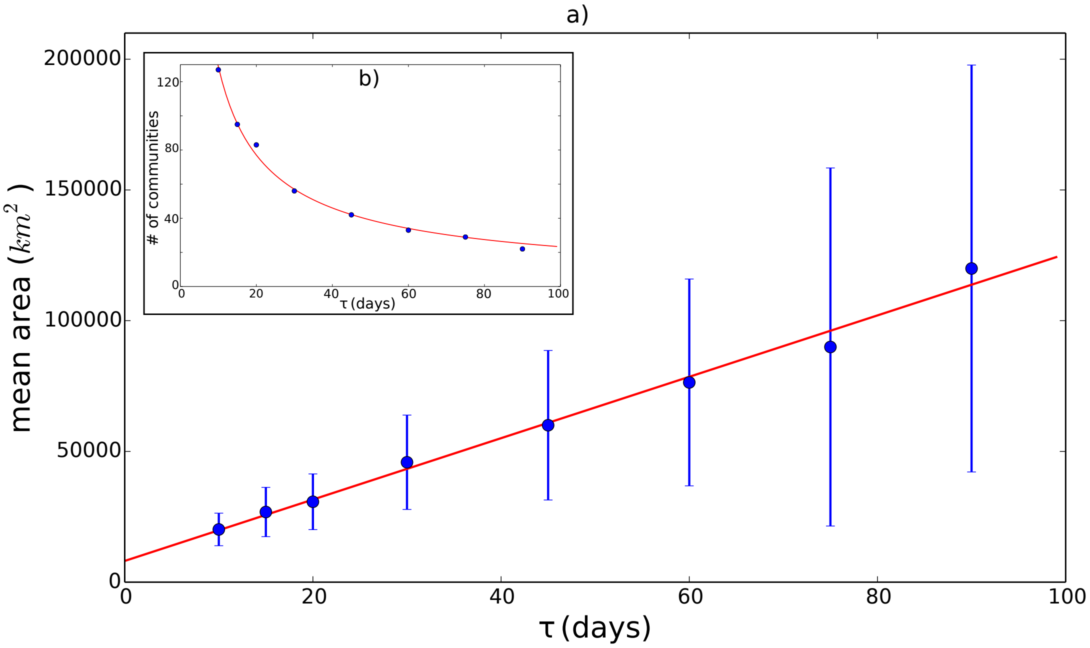

Figure 4.2: Panel a) shows the mean area (dots) of the communities detected by the *Infomap* algorithm for t_0 = July 1st 2011 as a function of τ. The straight line is a fit to the diffusive-growth-like relationship Area = 8109.6 + 1173.8 $* \tau$. The error bars indicate the standard deviation of the area distribution. Note the large dispersion in community sizes. The upper left inset b) shows the decay of the number of communities with τ.

some others appear and disappear when changing t_0. In order to identify persistent communities, i.e. those whose limits are relatively stable in space and time, we explore two averaging procedures leading to a mean -"climatological"- community partition. In a first approach we average a number of matrices $\mathbf{P}(t_0, \tau)$ corresponding to the same starting date (e.g. January 1st) for the ten different years of the data set (e.g. January 1st 2002, January 1st 2003, etc. until January 1st 2011). Figure 4.3 shows the *Infomap* partition of the network defined by the average matrix $\overline{\mathbf{P}(t_0, \tau)}$ made with the ten matrices $\mathbf{P}(t_0, \tau)$ using the same starting date for each of the 10 years (2002-2011). An example of t_0 in winter and another one in summer are displayed. The figure shows the most persistent communities for a particular month, averaging out the variability occurring over ten years. We remark than some communities have a rather small size (most of them reflecting shallow oceanic regions such as continental shelves), and that there is some inter-seasonal variability.

A second approach to obtain an average or climatological description of the community partition is illustrated in Fig. 4.4. Instead of applying only once *Infomap* on an averaged transport matrix, it is here applied 10 times separately on the 10 transport matrices corresponding to the same starting date for each of the 10 years (2002-2011). The color at a particular location of Fig. 4.4 indicates the frequency of occurrence (in these 10 partitions) at that location of a border between communities. Then, greener color indicates a more persistent community border. The strongest lines would represent true "barriers to transport" which remain fixed in space. Fuzzier lines may indicate intermittent border appearance, but also a larger wandering amplitude. Figures 4.4a and b display the situation in the same winter and summer days as in Fig. 4.3. Figure 4.4c shows a combination of them, equivalent to showing the barrier persistence sampled twice a year during the ten years.

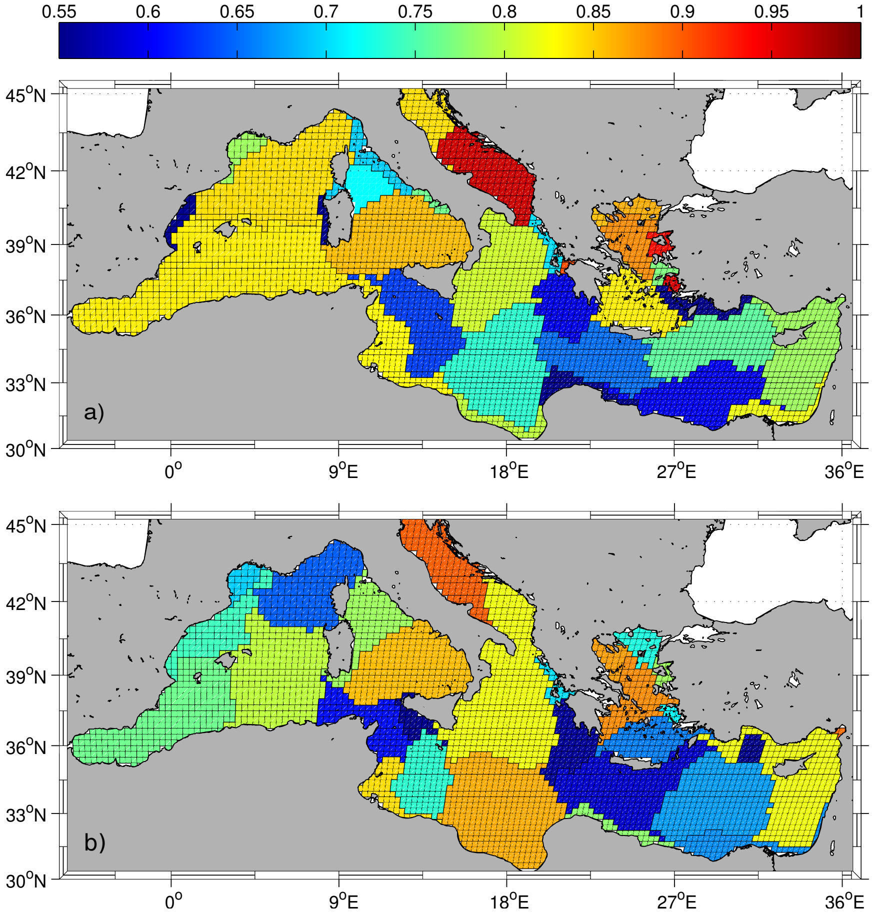

Figure 4.3: *Infomap* communities obtained from the average networks given by $\overline{\mathbf{P}(t_0,\tau)}$, with $\tau = 30$ days. Each community is colored by its coherence ratio. a) The average is over the 10 matrices corresponding to t_0 = January 1st in 10 years (2002-2011) of simulation; the number of communities is $p = 34$, the global coherence $\rho^{\tau}_{t_0}(\mathcal{P}) = 0.78$, and the global mixing $M^{\tau}_{t_0}(\mathcal{P}) = 0.68$. b) The average is over the 10 matrices corresponding to t_0 = July 1st in the 10 years 2002-2011; $p = 30$, $\rho^{\tau}_{t_0}(\mathcal{P}) = 0.77$, $M^{\tau}_{t_0}(\mathcal{P}) = 0.69$.

4.5

Comparison with spectral partitioning

Different methods based on the spectral properties of transport matrices have been previously used to identify and locate almost-invariant sets in flows (Froyland and Dellnitz, 2003; Froyland, 2005; Froyland et al., 2007; Dellnitz et al., 2009; Speetjens et al., 2013). They exploit the fact that for a set to remain almost invariant after the effect of the flow, it has to be related with eigenvectors of $\mathbf{P}(t_0,\tau)$ with eigenvalues close to

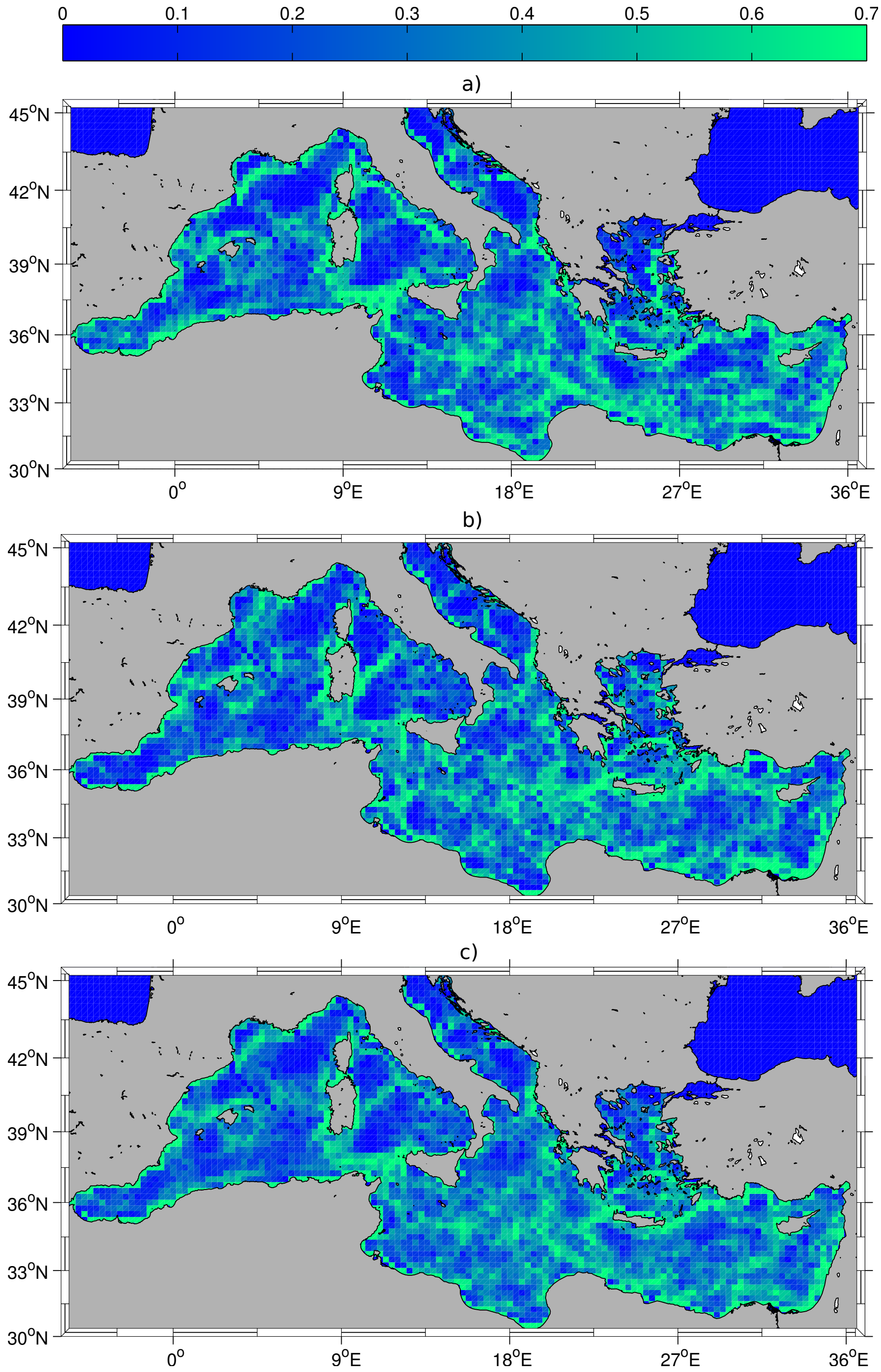

Figure 4.4: Persistence of community borders over time: Color code indicates the proportion of times one of the borders between communities has appeared at a given location. $\tau = 30$ days. a) t_0 = January 1st of (2002-2011). b) t_0 = July 1st of (2002-2011). c) The average of the two previous panels, eliminating the seasonal information.

1. Here we compare our partitioning obtained by *Infomap* with the one from those spectral methods. To be specific we consider the method described by (Froyland and Dellnitz, 2003). The technique in this last paper obtains a partition $\mathcal{P}$ minimizing in an approximate way the global coherence $\rho^{\tau}_{t_0}(\mathcal{P})$. To this end it computes eigenvectors associated to nearly vanishing eigenvalues of the Laplacian matrix (Newman, 2009) obtained from the symmetric part of $\mathbf{P}(t_0, \tau)$, and combines them using a fuzzy c-means clustering algorithm (Froyland and Dellnitz, 2003). Note that this approach eliminates any directionality information present in the transport network. Also, the c-means clustering can define as a single community pieces of the ocean which are geographically disjoint or in fact quite far apart, if this enhances the coherence defined in Eq. (4.4). In the method, one has to specify the number of eigenvectors being combined (we choose it to be 10) and the number of communities in the partition. Figure 4.5 shows the results

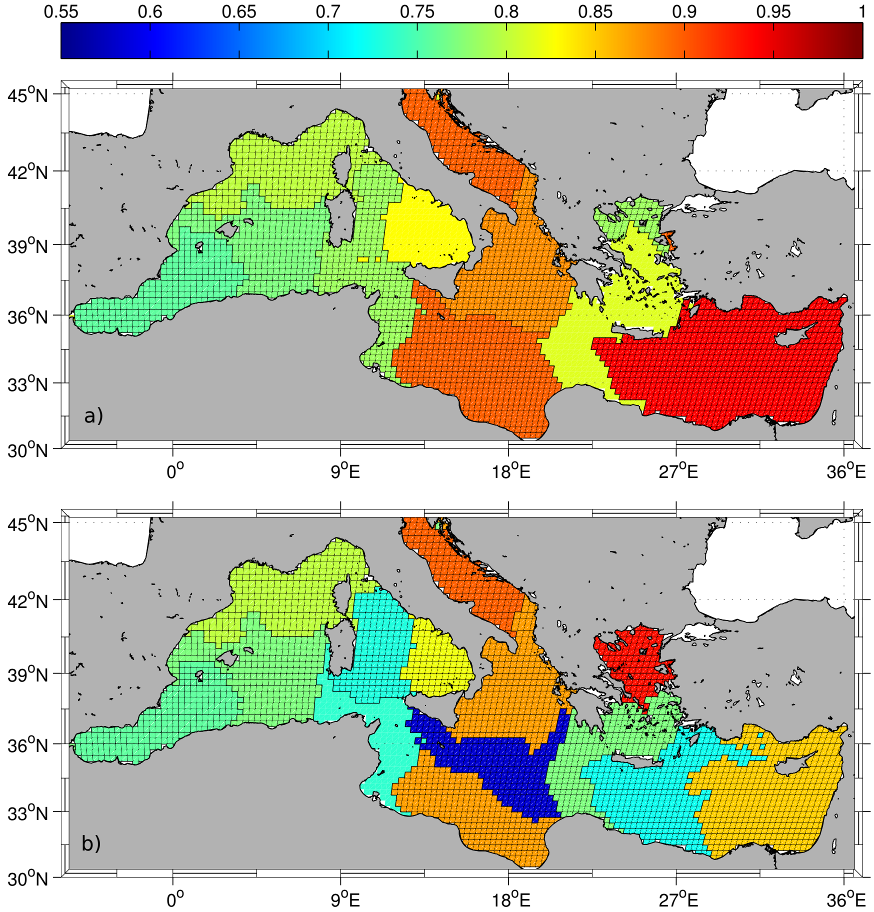

Figure 4.5: Community decomposition by the spectral method with fuzzy c-means clustering described in (Froyland and Dellnitz, 2003). The matrix used is the same average $\overline{\mathbf{P}(t_0, \tau)}$ as in Fig. 4.3 b), i.e. with t_0 = July 1st, averaged in the ten years 2002-2011, and $\tau = 30$ days. Ten eigenvalues are used. a) The number of communities is fixed to be $p = 10$; the global coherence is $\rho^{\tau}_{t_0}(\mathcal{P}) = 0.85$, and the global mixing is $M^{\tau}_{t_0}(\mathcal{P}) = 0.62$. In the Aegean, the southern yellow community is the only independent one: the portions of the Aegean further north are clustered by the c-means algorithm as being part of the same province as areas in the central Mediterranean with the same color. b) $p = 14$; $\rho^{\tau}_{t_0}(\mathcal{P}) = 0.78$, $M^{\tau}_{t_0}(\mathcal{P}) = 0.64$.

using the same average matrix $\overline{\mathbf{P}(t_0, \tau)}$ as in Fig. 4.3b, and imposing a partition in 10 and in 14 communities. The change in the number of communities leads to rearrangements in the Tyrrhenian, the central Mediterranean, the Aegean, and the Levantine basin. In panel a) some of the communities are made of disjoint pieces. Larger number of communities decreases the global coherence ratio (see caption of Fig. 4.5). If we try to increase the number of communities approaching the one given by *Infomap* we find that the clustering algorithm becomes unstable. Instabilities also occur when the number of links in the transport network becomes too high (as occurring for example when increasing τ beyond 1 month).

When compared with the *Infomap* decomposition we see that several of the boundaries coincide. But there are important differences, such as the wider range of community sizes and the sharper details revealed by *Infomap*. This is because a constraint of similar sizes for the communities associated to the same eigenvector needs to be imposed in the spectral method. When clustering several eigenvectors together this limitation is partially bypassed but still not removed. The values of the coherence ratio are of the same order or somehow larger for the spectral method, but note that the number of spectral communities has been kept much smaller to avoid the instabilities in the clustering algorithm. Since merging two communities into a single one increases the global coherence, joining some of the *Infomap* communities in Fig. 4.3 until arriving to 10 or 14 communities as in Fig. 4.5 would give rather large values of $\rho_{t_0}^{\tau}(\mathcal{P})$. As expected, the global mixing parameter is larger for the *Infomap* partition, but only by a small amount, reflecting that, even if internal mixing is not imposed in the spectral method, it is achieved to a reasonable extent.

From the methodological point of view, *Infomap* presents the advantage of determining itself the number of communities in the partition, whereas this needs to be fixed *a priori* (as well as the number of eigenvectors to be clustered) in the spectral approach. On the other hand, the spectral method is formulated as an algorithm to minimize the global coherence ratio, a quantity with a clear physical meaning. The quantity optimized by *Infomap* is a codeword length given in Eq. (1.20) of Section 1.2.3, an abstract information-theoretic object without a clear physical meaning. The heuristic interpretation of the optimization process leads to the 'large internal-small external connectivity' property for the communities, but a more rigorous understanding of the *Infomap* procedure is clearly needed (Rosvall and Bergstrom, 2008; Schaub et al., 2012).

The results of this section indicate that the *Infomap* methodology proposed here to identify coherent fluid regions seem more appropriate than spectral methods when a wide range of community sizes is expected, when internal mixing is a key parameter, or to minimize user input (such as entering the number of communities). Spectral methods seem appropriate when one is looking precisely for the sets defined mathematically as almost-invariant, the coherence ratio describes well the desired properties of the partition, and one expects a limited range of sizes.

4.6

Conclusions

To conclude, we described the basic steps to construct a flow network from fluid velocity data. We applied them to a surface flow field modeled for the Mediterranean sea. The resulting network is studied to characterize dispersion and mixing in different regions. One of the simplest network descriptors, the degree of a node, gives direct information on local stretching properties, classically associated to the finite-time Lyapunov exponents and their distributions. A family of network-entropy functions has been defined, aiming at describing higher-order statistical properties of fluid stretching (and then of dispersion and mixing) in terms of the network adjacency matrix. We find numerically that it provides a good estimation of the coarse-grained finite-time Lyapunov exponent.

We apply the network community detection method Infomap to identify coherent regions in the sea, well mixed internally but with little exchange among them. Our results indicate that the *Infomap* methodology proposed here to identify coherent fluid regions seems more appropriate than spectral methods when a wide range of community sizes is expected, when internal mixing is a key parameter, or to minimize user input (such as entering the number of communities). Spectral methods seem appropriate when one is looking precisely for the sets defined mathematically as almost-invariant, the coherence ratio describes well the desired properties of the partition, and one expects a limited range of sizes.

APPENDIX A

Relationship between network entropies and stretching statistics

In this Appendix we derive heuristically relationships between the network entropies and Lyapunov exponent statistics (in the two-dimensional case). Fig.A.1 illustrates the basic ideas. The assumptions are that dynamics is mainly hyperbolic in the region of interest, and that τ and the size Δ of the fluid boxes $\{B_i, i = 1, ..., N\}$ are such than the image of the boxes by the flow after a time τ are thin and long filaments. Boxes in the partition have been roughly aligned with expanding and contracting directions to make easier the heuristic arguments.

The point is to estimate the values of the matrix elements $\mathbf{P}(t_0, \tau)_{ik}$ given in Eq. (2.2):

$$\mathbf{P}(t_0, \tau)_{ik} = \frac{m\left(B_i \cap \Phi^{-\tau}_{t_0+\tau}(B_k)\right)}{m(B_i)} . \tag{A.1}$$

The quantity in the numerator of Eq. (A.1) is the area of the doubly-dashed thin filament in the left of Fig. A.1. If we assume that the forward FTLE $\lambda(\mathbf{x}_0, t_0, \tau)$ is approximately constant for $\mathbf{x}_0$ in this region, we have $m\left(B_i \cap \Phi^{-\tau}_{t_0+\tau}(B_k)\right) \approx \Delta^2 \exp(-\lambda_{ik}\tau)$ (see Fig. A.1), where λ_{ik} is this constant value. In consequence, $\mathbf{P}(t_0, \tau)_{ik} \approx \exp(-\lambda_{ik}\tau)$ if B_k is one of the boxes containing part of the image $\Phi^{\tau}_{t_0}(B_i)$ of B_i, and $\mathbf{P}(t_0, \tau)_{ik} = 0$ elsewhere.

Spatial features in typical forward FTLE fields are thin filaments with nearly constant value λ. They are elongated along the expanding directions (Haller, 2001; Shadden et al., 2005) and have widths of the order of $l \exp(-\lambda\tau)$, where l is the size of the velocity field inhomogeneities, i.e. the size of the Eulerian structures driving the flow. Then, the uniformity condition we are imposing is $\Delta < l$, i.e. discretization boxes smaller than Eulerian structures. In our Mediterranean example, Δ is smaller than the dominant mesoscale structures in the sea, but some of the smaller features in the velocity field can have some impact on the validity of the uniformity condition.

We can use our estimation of $\mathbf{P}(t_0, \tau)_{ik}$ to compute the sum appearing in the network entropies definition Eq. (3.7). The assumption of uniform FTLE inside region $B_i \cap$

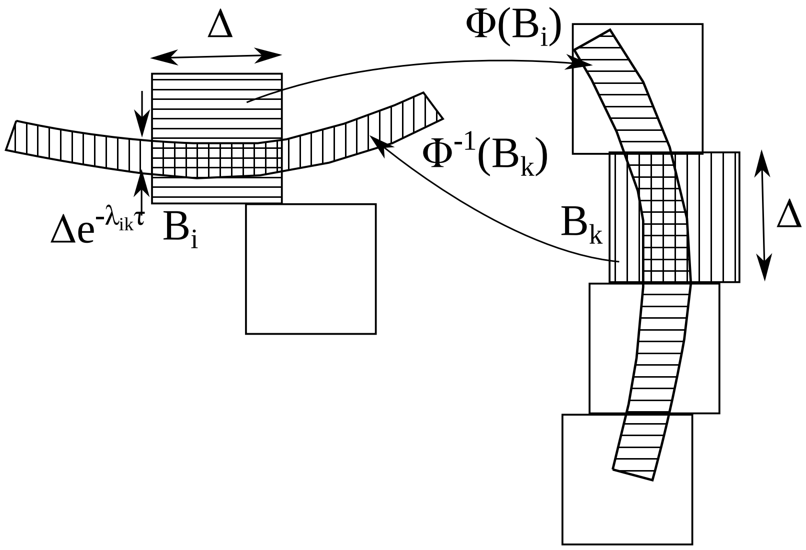

Figure A.1: Schematics of the stretching (forward and backwards in time) of fluid boxes of sidelength Δ corresponding to network nodes. $\Phi(B_i)$ is a shortcut for $\Phi^{\tau}_{t_0}(B_i)$, and $\Phi^{-1}(B_k)$ is a shortcut for $\Phi^{-\tau}_{t_0+\tau}(B_k)$. λ_{ik} is the value of the forward FTLE $\lambda(\mathbf{x}_0, t_0, \tau)$ in the doubly dashed region $B_i \cap \Phi^{-1}(B_k)$.

$\Phi^{-\tau}_{t_0+\tau}(B_k)$ allows us to freely replace functions of λ_{ik} by average values in that region:

$$\begin{aligned}
&\sum_{k=1}^{N} (\mathbf{P}(t_0,\tau)_{ik})^q \approx \sum_{k=1}^{N} e^{-q\lambda_{ik}\tau} \approx , \\
&\sum_{k=1}^{N} \frac{1}{\Delta^2 e^{-\lambda_{ik}\tau}} \int_{B_i \cap \Phi^{-\tau}_{t_0+\tau}(B_k)} e^{-q\lambda(\mathbf{x}_0,t_0,\tau)\tau} d\mathbf{x}_0 \approx , \\
&\sum_{k=1}^{N} \frac{1}{\Delta^2} \int_{B_i \cap \Phi^{-\tau}_{t_0+\tau}(B_k)} e^{(1-q)\lambda(\mathbf{x}_0,t_0,\tau)\tau} d\mathbf{x}_0 = , \\
&\qquad \frac{1}{\Delta^2} \int_{B_i} e^{(1-q)\lambda(\mathbf{x}_0,t_0,\tau)\tau} d\mathbf{x}_0 \, .
\end{aligned} \tag{A.2}$$

Which, using definition (3.7), implies:

$$e^{(1-q)\tau H_i^q(t_0,\tau)} \approx \left\langle e^{(1-q)\tau\lambda(\mathbf{x}_0,t_0,\tau)} \right\rangle_{B_i} . \tag{A.3}$$

This is the sought relationship between network entropies and moments of the stretching factor $e^{\lambda\tau}$. For $q = 0$ we reobtain Eq. (3.6). In the limit $q \to 1$ we get $H_i^1(t_0,\tau) \approx \langle \lambda(\mathbf{x}_0, t_0, \tau) \rangle_{B_i} = \lambda_i(t_0, \tau)$. The arguments above can be repeated to get the same relationship (A.3) between network entropies in the backwards time direction and backwards Lyapunov exponents.

All these expressions are similar to the ones presented for example by (Paladin and Vulpiani, 1987) relating Rényi entropies and generalized Lyapunov exponents defined from moments of the stretching factor $e^{\lambda\tau}$. But here the moments are not by averaging along a dynamic trajectory but inside a box B_i. In the same way as the value of any of the network entropies at node i characterizes the inhomogeneity in the fluxes sent from i to other nodes, the difference between the different entropies (different q) at a single node i characterizes the inhomogeneity of the FTLE inside box B_i. This is a way by which small-scale features present in the Lagrangian trajectories get statistically represented in the network description. Relationships such as (A.3) are not exact for finite Δ and τ, but we expect them to become more accurate for increasing τ and decreasing Δ.

APPENDIX B

Computing network entropies with nodes of differing area

We consider the problem of defining a family of network entropies where the node's area could vary arbitrarily. We start studying the motion of an initial concentration of tracer in a box i assuming that after a sufficient long time it will evolve into a long and thin filament. We can define the length of this filament as:

$$\bar{L} = \frac{1}{\Delta_i} \int d\mathbf{x}_0 e^{\tau\lambda(\mathbf{x}_0, t_0, \tau)} \,. \tag{B.1}$$

Now we are interested in relating the number of nodes "touched" by the filament with the real length of the filament (assuming a dynamics dominated by strain without folding of the filaments). We can write:

$$K_O(i) \approx \frac{\bar{L}}{\Delta^{ef}} \,, \tag{B.2}$$

where $K_O(i)$ is the out-degree of the node i and Δ^{ef} is the effective node's size that should be understood like the mean size of the boxes crossed by the filament. We find an exact expression for it:

$$\Delta^{ef} = \frac{\sum_j^N \Delta_j \delta(\mathbf{P}(t_0, \tau)_{ij})}{K_O(i)} \,. \tag{B.3}$$

Where $\mathbf{P}(t_0, \tau)_{ij}$ is the element ij of the transport matrix and $\delta(\mathbf{P}(t_0, \tau)_{ij})$ is zero if $\mathbf{P}(t_0, \tau)_{ij} = 0$ and equal to one otherwise. Using the last definition in the previous equation we find:

$$K_O(i) \approx \frac{K_O(i)}{\sum_j^N \Delta_j \delta(\mathbf{P}(t_0, \tau)_{ij})} \bar{L} \,. \tag{B.4}$$

From this, for consistence, we find the relation:

$$\bar{L} = \frac{1}{\Delta_i} \int d\mathbf{x}_0 e^{\tau\lambda(\mathbf{x}_0, t_0, \tau)} \approx \sum_j^N \Delta_j \delta(\mathbf{P}(t_0, \tau)_{ij}) \,, \tag{B.5}$$

dividing by Δ_i we finally have:

$$\frac{1}{\Delta_i^2} \int d\mathbf{x}_0 e^{\tau\lambda(\mathbf{x}_0, t_0, \tau)} = \langle e^{\tau\lambda(\mathbf{x}_0, t_0, \tau)} \rangle_{B_i} \approx \sum_j^N \frac{\Delta_j}{\Delta_i} \delta(\mathbf{P}(t_0, \tau)_{ij}) \,. \tag{B.6}$$

APPENDIX B. COMPUTING NETWORK ENTROPIES WITH NODES OF DIFFERING AREA

We define now a generalized network entropies family in this way:

$$G_i^q(t_0,\tau) = \frac{1}{(1-q)\tau}\log\sum_j^N\left[\frac{\Delta_j}{\Delta_i}\Big(\frac{\Delta_i}{\Delta_j}\mathbf{P}(t_0,\tau)_{ij}\Big)^q\right]. \tag{B.7}$$

We take now the limit $q \to 0$ of G_i^q and we find:

$$G_i^0(t_0,\tau) = \frac{1}{\tau}\log\sum_j^N\left[\frac{\Delta_j}{\Delta_i}\delta(\mathbf{P}(t_0,\tau)_{ij})\right] \approx \lambda(\mathbf{x}_0,t_0,\tau)\,. \tag{B.8}$$

That match exactly with Eq. B.6. Taking instead the limit $q \to 1$ we have (using l'Hôpital's rule and the normalization constraint $\sum_j^N \mathbf{P}(t_0,\tau)_{ij} = 1$):

$$G_i^1 = -\frac{1}{\tau}\left[\sum_j^N\left[\mathbf{P}(t_0,\tau)_{ij}\log\frac{\mathbf{P}(t_0,\tau)_{ij}}{\Delta_j}\right] + \log(\Delta_i)\right]. \tag{B.9}$$

That match exactly with the definition of Finite Time Entropy for an uneven discretization grid given by (Froyland and Padberg-Gehle, 2012).

Part II

Ecological applications

CHAPTER 5

Marine Protected Areas (MPAs) and oceanic ecological connectivity

5.1 Introduction

Oceanic ecosystems are impacted by multiple human-induced stressors, including habitat destruction, pollution, overfishing and global climate change. Marine protected areas (MPAs), used for the management and conservation of marine ecosystems, are considered effective to mitigate some of these impacts (Lester et al., 2009). Successful MPA design is however complicated primarily due to the difficulties in quantifying the movements of organisms, especially at larval stage (Shanks, 2009), in resolving the multi-scale variability of ocean currents (Siegel et al., 2008) and in apprehending the spatial scales and biogeography of the seascape (Hamilton et al., 2010).

Marine population connectivity, i.e. the exchange of individuals among geographically separated subpopulations, depends on numerous factors including spawning outputs, larval dispersal, habitat availability, trophic interactions and adult movements (Cowen and Sponaugle, 2009; Game et al., 2009). Among them, larval dispersal has been identified as a crucial factor for structuring oceanic populations (Cowen et al., 2006) and for determining broad-scale ecological connectivity (Treml et al., 2012). It also plays a major role in assuring population persistence in a MPA network, especially when target species show long-distance dispersal (Shanks, 2009). As such, patterns and magnitude of larval connectivity have been used to design MPAs (Lester et al., 2009) and assess their efficiency (Pelc et al., 2010). This chapter focuses on the dispersion of larvae by ocean currents at basin-scale in the Mediterranean Sea, assuming they are passively transported by the flow (i.e. neglecting larval behavior), to inform the design of marine reserves.

Many biophysical modeling studies (Cowen et al., 2006; Siegel et al., 2008), including Lagrangian approaches, examined marine connectivity from the so-called "connectivity matrix" which represents the probability of larval exchange between distant sites. Previous analyses were mainly limited to the strengths of pair-wise connections, i.e. the links from one coastal site, or MPA, to another distant one (Corell et al., 2012; Vaz et al., 2013). Another perspective to investigate connectivity is the analysis of dispersal network topologies (Treml et al., 2012; Kool et al., 2013). Recent studies applied tools derived from Graph Theory to document regional connectivity of near-shore MPAs in the Baltic Sea (Nilsson-Jacobi et al., 2012), the Mediterranean Sea (Andrello et al., 2013) and in the Great Barrier Reef region (Thomas et al., 2014). While our understanding of connectivity at small- and regional-scales has improved, previous efforts focused mainly on coastal/insular areas and did not provide a characterization of the seascape connectivity.

The significance of this shortcoming is emphasized by the growing interests for the implementation of MPAs in the pelagic ocean (Pala, 2013; Guidetti et al., 2013) which also shelters biodiversity and important ecological processes (Game et al., 2009; Kaplan et al., 2010a). Designing open-ocean MPAs is challenging partly because larval connectivity and pelagic habitats are difficult to assess in such vast and dynamic environment.

We stress that the need for an ecosystem-based management of marine resources has been emerging (Pikitch et al., 2004; Kaplan et al., 2010b), especially in the Mediterranean basin (Coll et al., 2012, 2013; Guidetti et al., 2013). For instance, recent studies considered an ecosystem-based approach to optimize the location of reserves for several species based on multi-factorial analysis (Lagabrielle et al., 2012) or to inform MPA efficiency by modelling trophic interactions within the whole ecosystem (Colleter et al., 2012). In the context of assessing larval connectivity for MPAs design, it implies the need of considering the ecosystem as a whole rather than focusing on a specific organism. This is implemented here by studying a range of different Pelagic Larval Durations (PLDs) and periods of spawning which are, under certain assumptions, biologically relevant for a number of Mediterranean species.

A compilation of the mean PLDs of 62 littoral Mediterranean fishes revealed they span 10-70 days depending on the species considered, with large intra-species variability (Macpherson and Raventos, 2006). Considering the ”basin-scale” angle of our study, we focus on species with a wide geographical range and potential for large-distance dispersal (see Tab. 5.1). These organisms are usually characterized by pelagic spawning, long PLDs (≥ 20 days) and offshore larval distribution (although many combinations of such early life traits exist). We retain a PLD of 30 days which is the best estimate available for a few iconic species of the Mediterranean ecosystem, including some demersal fishes (e.g. the groper *Epinephelus marginatus*, the blenny *Lipophrys canevai*, the wrasse *Labrus viridis*, the goatfish *Mullus surmuletus*, the bream *Sarpa salpa*) and invertebrates (e.g. the crab *Pachygrapsus marmoratus*). We also considered a PLD of 60 days since other Mediterranean fishes (e.g. the blenny *Lipophrys trigloides*) and most marine invertebrates (echinoderms like the sea-star *Astropecten aranciacus*, some molluscs and many exploited crustaceans) are characterized by long PLD.

Dispersal potential can also be influenced by other mechanisms than PLD, such as early life traits. However, because the precise description of spawning strategy and larval distribution of marine organisms remains elusive, a classification was proposed based on its preferential season and location of occurrence. A large majority of the species studied (e.g. *Epinephelus marginatus, Lipophrys canevai, Mullus surmuletus*) spawn in late spring / early summer, so we assume that summer is the season of their planktonic life. Some others (e.g. *Sarpa salpa, Lipophrys trigloides*) prefer a late autumn / early winter spawning, resulting in the highest abundance of larvae observed in winter. Concerning spatial preferences, the ”coastal spawners” (e.g. *Labrus viridis*) release their eggs close to the bottom in shallow areas and then their planktonic larvae are concentrated in the coastal ocean. In contrast, the ”pelagic spawners” (e.g. *Epinephelus* *marginatus, Mullus surmuletus, Sarpa salpa*) spawn in the open-ocean with their larvae found widespread offshore (Macpherson and Raventos, 2006).

Marine species	Taxonomy	Category (adults habitats)	Larval distribution	Estimated PLD (days)	Larval season (following spawning)
Blenny *Lipophrys trigloides*	Vertebrate, Fish	Littoral demersal (benthic)	Inshore	67	Winter
Blenny *Lipophrys canevai*	Vertebrate, Fish	Littoral demersal (benthic)	Offshore	30	Summer
Rainbow Wrasse *Coris julis*	Vertebrate, Fish	Littoral demersal (benthopelagic)	Offshore	21-34	Summer
Green Wrasse *Labrus viridis*	Vertebrate, Fish	Littoral demersal (benthopelagic)	Inshore	31-34	Spring/ Summer
Goat Fish *Mullus surmuletus*	Vertebrate, Fish	Littoral demersal (benthopelagic)	Offshore	30	Spring/ Summer
Dusky Groper *Epinephelus marginatus*	Vertebrate, Fish	Littoral demersal (benthopelagic)	Offshore	25-30	Summer
Salema Porgy *Sarpa salpa*	Vertebrate, Fish	Littoral demersal (benthopelagic)	Offshore	32	Winter
Shore Rockling *Gaidropsarus mediterraneus*	Vertebrate, Fish	Littoral demersal (benthic)	Offshore	43	Winter
Two-banded Seabream *Diplodus vulgaris*	Vertebrate, Fish	Littoral/Shelf demersal (benthopelagic)	Offshore	29-58	Winter
White Seabream *Diplodus sargus*	Vertebrate, Fish	Littoral demersal (benthopelagic)	Inshore	28	Winter
Gilthead Seabream *Sparus aurata*	Vertebrate, Fish	Littoral/Shelf demersal (benthopelagic)	Offshore	40-50	Winter
Bullet Tuna *Auxis rochei*	Vertebrate, Fish	Shelf pelagic (epipelagic)	Offshore	16	Spring/ Summer
Sandsmelt Fish *Atherina spp.*	Vertebrate, Fish	Littoral pelagic (epipelagic)	Inshore	9-15	Spring/ Summer
Dolphin Fish *Coryphaena hippurus*	Vertebrate, Fish	Shelf pelagic (epipelagic)	Offshore	?	Spring/ Summer
European Anchovy *Engraulis encrasicolus*	Vertebrate, Fish	Oceanic pelagic (epipelagic)	Offshore	37	Summer
Bluefin Tuna *Thunnus thunnus*	Vertebrate, Fish	Oceanic pelagic (epipelagic)	Offshore	30	Summer
Ray Bream *Brama brama*	Vertebrate, Fish	Oceanic pelagic (epipelagic)	Offshore	?	Summer
Gilt Sardine *Sardinella aurita*	Vertebrate, Fish	Oceanic pelagic (epipelagic)	Offshore	60	Summer
European Hake *Merluccius merluccius*	Vertebrate, Fish	Shelf/Oceanic demersal (benthopelagic)	Offshore	40-60	Summer/ Autumn
Horse Mackerel *Trachurus mediterraneus*	Vertebrate, Fish	Shelf/Oceanic pelagic (epipelagic)	Offshore	?	Summer
European Seabass *Dicentrarchus labrax*	Vertebrate, Fish	Littoral/Shelf demersal (benthopelagic)	Offshore	40	Winter
Sea Star *Astropecten aranciacus*	Invertebrate, Echinoderms	Littoral demersal (benthic)	Inshore	60	Spring/ Summer
Marbled Crab *Pachygraptus marmoratus*	Invertebrate, Crustaceans	Littoral/Shelf demersal (benthic)	Inshore	30	Spring/ Summer
Other crustaceans (e.g. Lobster)	Invertebrate, Crustaceans	Littoral/Shelf demersal (benthic)	Variable	~30-300	Variable
Other molluscs (e.g. Oyster)	Invertebrate, Molluscs	Littoral demersal (benthic)	Variable	~10-100	Variable

Table 5.1: Literature review of biological traits for some emblematic Mediterranean marine species with wide geographical range and potential for large-scale dispersal. With our parameter values and under our assumptions, the broad-scale connectivity patterns evidenced in this paper are applicable for those organisms that belong to different trophic levels of the Mediterranean food webs. Note that the vertical positioning of eggs/larvae may vary but for simplicity, they are assumed here to remain in the $\sim 50-150$ m thick Mixed Layer Depth, thus seeing relatively similar transport patterns. While considering the ecosystem as a whole (rather than focusing on a specific organism) for an ecosystem-based management of marine resources is now widely accepted, one could obtain refined analyses by tuning our modeling framework to any target specie with well-known biological traits and to any oceanic area. Category: Littoral $\sim 0-50$ m; Shelf $\sim 50-200$ m; Oceanic ≥ 200 m. Larval distribution: Inshore $\sim 0-50$ m; Offshore ≥ 50 m (i.e. shelf and oceanic waters).

5.2

Lagrangian bio-physical modeling for the Mediterranean

The Lagrangian approach is a natural perspective to characterize transport phenomena affecting free-swimming larvae (Corell et al., 2012; Vaz et al., 2013). We use the method described in Part I to simulate Lagrangian trajectories and build connectivity matrices as adjacency matrix of flow networks.

We focus again on the upper-ocean dynamics over years 2002−2011 with the use of daily horizontal flow fields at 8 m depth, representing the surface mixed layer in which larvae are assumed to be homogeneously distributed. Lagrangian particles are dispersed as two-dimensional passive drifters. Note that due to the non-fully incompressible horizontal flow field, vertical velocities may become significant in regions of strong divergence (e.g. coastal upwelling) and convergence (e.g. deep water formation). Neglecting vertical movements is however a reasonable assumption here because most particles remain in the selected layer over short time-scales ($\leq$ 2 months) since horizontal velocities are several orders of magnitude higher than vertical ones (d'Ovidio et al., 2004). Another simplification is the passive character of the particles, the implementation of more complex larval behavior (e.g. vertical migration, mortality, settlement) being envisaged for future work. Under these assumptions, larval dispersal is modulated by the PLD, the period of spawning and the time-varying oceanic circulation.

Initial (t_0) and integration ($\tau \sim$ PLD) times are chosen according to the typical biological traits of marine organisms. Given the limited knowledge of their life cycles (Shanks, 2009), we investigate basin-scale larval connectivity from an ecosystem-based approach (Coll et al., 2012; Guidetti et al., 2013), rather than focusing on a particular target species. To do so, we retain two different PLDs ($\tau = 30, 60$ days) and both winter and summer spawning by considering three successive starting times t_0 for each season ($t_0 = 1^{st}, 15^{th}$ and 31^{st} of January and $t_0 = 1^{st}, 15^{th}$ and 31^{st} of July) to account for episodic and variable spawning events (Macpherson and Raventos, 2006; Shanks, 2009; Andrello et al., 2013) over the years 2002 − 2011. These modeling choices are ecologically meaningful for a number of Mediterranean organisms, especially those with wide geographical range and potential for large-distance dispersal. Sensitivity of our results to the parameter τ was tested by performing additional simulations for $\tau = 45$ days for winter ($t_0 = 1^{st}$ Jan.) and summer ($t_0 = 1^{st}$ Jul.). A total of 140 factorial experiments (with starting times covering 2 seasons over 10 years, for 3 PLDs) allow the construction of 140 flow connectivity matrices from which robust spatial patterns of larval connectivity in the entire Mediterranean basin can be extracted.

5.3

Oceanographic characterization of hydrodynamic provinces

We apply the method described in Part I to extract communities from each constructed network and we interpret them as time-dependent hydrodynamical provinces in the sea. To evaluate the significance of the network partitioning, we use the coherence ratio associated with each province ρ_k defined in Eq. 4.2. We perform this analysis retaining three different PLDs ($\tau = 30, 45, 60$ days) and consider winter ($t_0 = 1^{st}$ Jan.) and summer

($t_0 = 1^{st}$ Jul.) spawning over the years 2002 – 2011. A total of 60 factorial (with starting times covering 2 seasons over 10 years, for each of the 3 PLDs) allow the construction of 60 connectivity matrices from which hydrodynamical provinces are extracted (Rossi et al., 2014).

The provinces and their boundaries are dynamical objects that evolve in space and time with different dimensions, shapes and locations (e.g. Fig. 5.1), due to the important variability of the ocean circulation (Millot and Taupier-Letage, 2005). The method captures an elevated number of communities in the network, with 65 provinces using a PLD = 30 days and only 32 for PLD = 60 days on the exemplary calculations displayed on Fig. 5.1. Intuitively, the longer the tracking time, the lower the number of provinces detected and the larger their mean area. On average over the ensemble of experiments, community detection results in 61, 46, 36 provinces characterized by a mean area of 4.12×10^4, 5.5×10^4, 6.8×10^4 (in km^2) for $\tau = 30, 45, 60$ days, respectively. Because of the time-varying flow (Siegel et al., 2008), both release time and tracking duration (simulating respectively the initiation and duration of the pelagic larval phase) affect the spatial partitioning.

Most province boundaries match very well the mean flow streamlines (Fig. 5.1), suggesting high oceanographic relevance. While isolated streamlines are found in the cores of provinces, dense ones usually coincide with the detected boundaries. Hydrodynamical provinces are delimited by intense oceanic mesoscale structures such as jets, meanders, fronts and eddies. These features, which influence the topology of the transport network and thus the community detection, were recently reported to strongly impact connectivity (Vaz et al., 2013). For instance, some mesoscale eddies are extracted as quasi-circular single provinces (e.g. in the Alboran Sea and in the southern Levantine basin), in good agreement with the flow streamlines (Fig. 5.1a). Other mesoscale structures are contained in larger provinces. The method allows the optimal detection of coherent oceanic sub-regions originating from the ocean circulation and its multiscale variability.

The coherence ratios are generally elevated ($\rho \in [0.5, 1]$) and variable (Fig. 5.1). Although it depends on both the local leaking processes and the area of a given province, there is no apparent relationship between the size of the sub-region and its coherence ratio. Overall, $\rho \geq 0.8$ are often seen in the Aegean and Adriatic Seas. The Alboran, Balearic, Tyrrhenian, and Adriatic Seas are characterized by relatively large provinces, whereas the Levantine, Aegean and south Ionian and Libyan Seas are subdivided in rather small ones. Note also that some provinces are composed of non-contiguous boxes. This occurs especially within the pathways of fast-flowing currents as the Algerian Current, the Atlantic-Ionian stream (south Ionian, Libyan and south-east Levantine sea) and the Liguro-Provencal Current (Ligurian sector).

We focus now on larval dispersal patterns examining the frequency of occurrence of province boundaries (as explained in Part I) across the ensemble of experiments to identify recurrent frontal systems and relatively stable hydrodynamical unit. Over most coastal/shallow regions, boundaries occur in various locations and orientations, resulting in no apparent structure (dark red patches in Fig. 5.2). These disorganized patterns characterize oceanic environments with complex circulation in which spatial-scales of connectivity are highly variable (Siegel et al., 2008). They are observed in most insular regions (Balearic, Tuscan and Aegean archipelagos, Corsica, Sardinia, Crete,

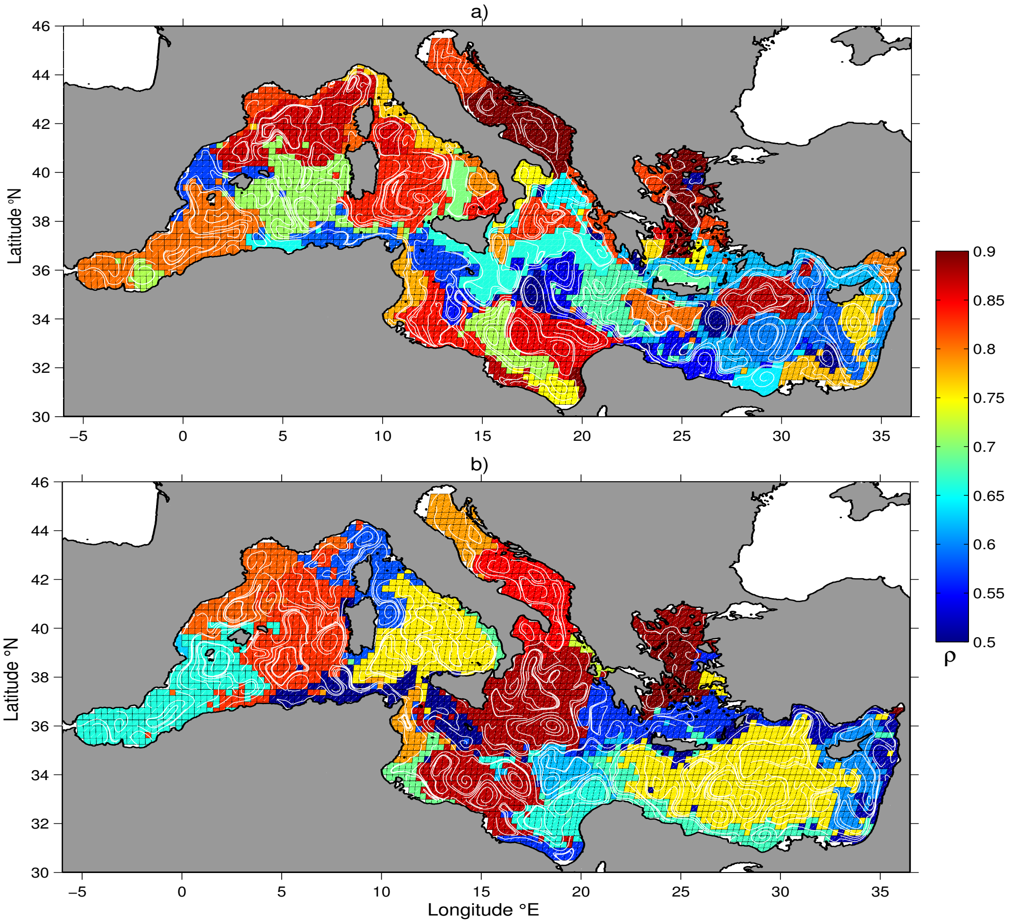

Figure 5.1: Hydrodynamical provinces extracted from the connectivity matrices of a) Winter 2011 (t_0 = 1^{st} Jan.) using $\tau \simeq PLD = 30$ days and b) Summer 2011 (t_0 = 1^{st} Jul.) using $\tau \simeq PLD = 60$ days. Each province is colored according to its ρ value (ranging from 0.5 to 0.9). As already showed in Chapter 4, the number of provinces decreases with τ. White streamlines represent the simulated flow averaged over the period of integration (i.e. a) 1^{st}-30^{th}Jan. 2011 and b) 1^{st}Jul.-29^{th}Aug. 2011) and they match very well with province boundaries.

Cyprus), in the Tunisian-Sicilian strait (also punctuated by small islands) and over narrow continental shelves (Italian, French, Catalan, Libyan-Egyptian and Israelian-Lebanon shelves) (Fig. 5.2a). In contrast, wide continental shelves are organized as coherent hydrodynamical units whose offshore limits match the 200 m isobath. The gulf of Lion is delimited by a frontier coinciding with the Catalan front and associated Northern current (Bouffard et al., 2010) (an extension of the Liguro-Provencal current) . For long PLDs only, the Tunisian-Libyan shelf appears as two units in summer (Fig. 5.2b), merging into a single one in winter. The oceanic frontiers constituted by such currents/fronts are likely to prevent coastal larvae from escaping wide continental shelves.

In the open ocean, clear hydrodynamical units emerge (Fig. 5.2), organized as large "gyre" systems with rare occurrence of boundaries (white/yellow colors) in their center and semi-persistent frontiers (light/dark red colors) aligned along their perimeters. Elevated connectivity within each subdivision but little exchange between them are expected, thus providing basin-scale patterns of larval dispersal. Large hydrodynamical units are found in the western Mediterranean basin, the Adriatic sea, the Tyrrhenian sea (Fig. 5.2a and b) and only at longer time-scales in the north-Ionian and Levantine seas (Fig. 5.2b). Most of these open-ocean frontiers are located along well-known oceanographic features (Millot and Taupier-Letage, 2005), some of them recognized as partial transport barriers. For instance, the so-called Oran-Almeria front separates the Alboran sea from the rest of the Mediterranean Sea. It appears here rather extending from Oran to Cartagena, some 200 km away than previously documented. The Balearic front is another semi-permament transport barrier (Mancho et al., 2008) passing north of the Balearic archipelago in the Balearic current (Bouffard et al., 2010) and elongating eastward in winter. North of this quasi-zonal boundary, a large hydrodynamical unit composed of the Lion gyre and Ligurian sea is separated from the Balearic sea at short time-scale. The Tyrrhenian sea is consistently organized as a two-gyre system using both PLDs. For the 30-day integration the Adriatic sea is subdivided by bathymetric gradients ($\sim$ 100 and 200 m isobaths, i.e. off the Gargano promontory) into a northern, central and southern Adriatic gyres, the two latter units merging for PLDs of 60 days.

Surprisingly, some open-ocean areas, such as the Ionian, Levantine and Aegean basins (Fig. 5.2a), are characterized by disorganized dispersal patterns and stochastic larval connectivity (Siegel et al., 2008). They become more structured at longer time-scales with the emergence of the Western Ionian gyre, the Shikmona gyre and a large system encompassing the Rhodes, Ierapetra and Mersa-Matruh gyres (Millot and Taupier-Letage, 2005). The eastern Aegean sea has disorganized dispersal patterns whereas small hydrodynamical units appear in its northern and western parts, in good agreement with its thermal structure (Poulos et al., 1997).

More generally, regions with no apparent spatial patterns at short PLDs see the emergence of spatial structures for longer integration time. Oceanic areas already identified as gyral systems for short time-scales have their diameter increasing with the integration time, ultimately merging with their neighbors.

Note that most of these hydrodynamical units are quite consistent with the trophic clusters obtained from satellite chlorophyll data (d'Ortenzio and d'Alcalá, 2009), suggesting they also delimits specific biogeochemical provinces (Longhurst, 2006). Indeed, although this study focuses on passive larvae, the unveiling of well-known oceanic fronts and gyres hint that the spatial distribution of other tracers (e.g. salinity, temperature, chlorophyll-*a*, dissolved nutrients) are also influenced by similar transport patterns.

5.4

Implications for the design of MPAs

The geographical structure of larval dispersal in the seascape influences largely the connectivity of marine reserves. The MPAs located within large and stable hydrodynamical units (Fig. 5.2) are interconnected, in good agreement with (Andrello et al., 2013) who identified similar MPA clusters in the Algerian, Balearic, Adriatic and Tyrrhenian seas, respectively. Further information is obtained with the analysis of three complementary proxies of connectivity defined as followed. We analyze the mean spatial scales of

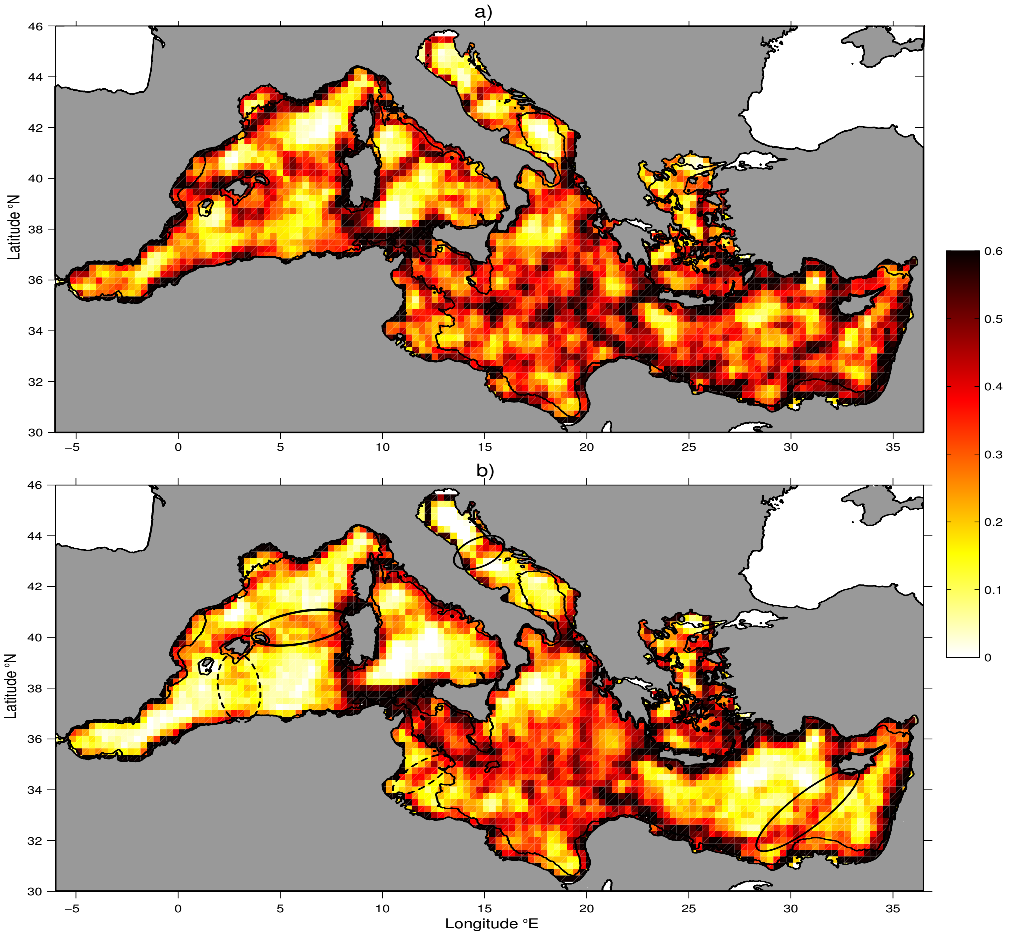

Figure 5.2: Frequency of occurrence of province boundaries at each ocean node from the ensemble of 20 experiments (all winters/summers of 2002 – 2011) for each PLD: a) $\tau \simeq PLD = 30$ days and b) $\tau \simeq PLD = 60$ days. While in Chapter 4 we focused only on $\tau = 30$ here we show differences among different PLDs. Black ellipses in b) highlight the frontiers which have significant seasonality: plain ellipses indicate a preferential occurrence in winter and dotted ellipses in summer. Black contours represent the 200 m isobath.

larval dispersal (Fig. 5.3a) and the mean local coherence (inversely related to leaking, Fig. 5.3b) by averaging over the ensemble of experiments the area and the coherence ρ, respectively, of the time-dependent province encompassing each MPA. While these two metrics are solely influenced by the flow, the mean number of interconnected MPAs (i.e. temporally averaged number of MPAs encountered within the same time-dependent province, Fig. 5.3c) depends also on the density of existing reserves.

Larval connectivity and dispersal potentials are highly variable among the Mediterranean MPAs (Fig. 5.3). Reserves in the Adriatic and Aegean seas are characterized by small dispersal surface ($\leq 5 \times 10^4\ km^2$) and among the highest coherence ($\rho \geq 0.8$). This suggests a low connectivity which is also reflected in the few interconnected MPAs (≤ 8) despite their relatively high density. MPAs located around isolated islands are associated with modest dispersal surface ($\sim 4 - 8 \times 10^4\ km^2$) and low coherence ($\rho \leq 0.7$). Typical of these insular environments (Vaz et al., 2013), complex circulation patterns (islands' wake, eddies, retention...) result in a moderate connectivity and high tem-

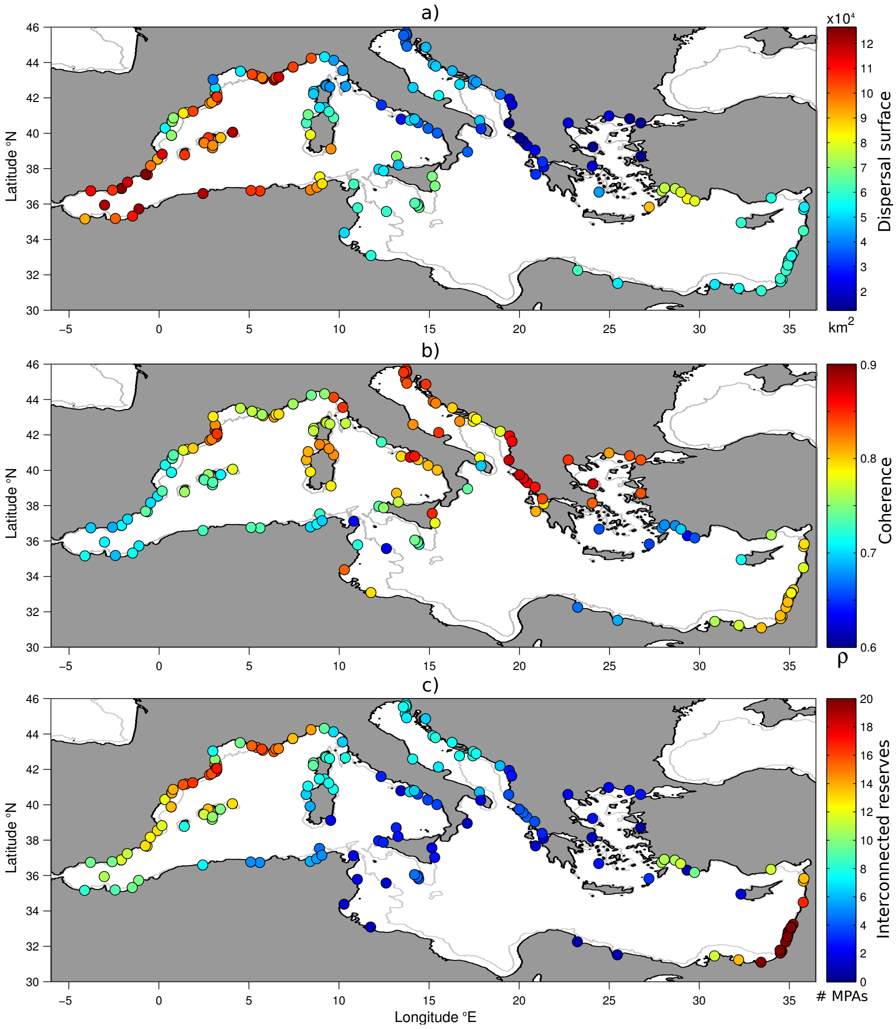

Figure 5.3: Spatial variability of MPAs connectivity derived from three complementary metrics averaged over all winter/summer experiments over 2002 – 2011 using $\tau \simeq PLD = 30$ days. a) Mean area (in km^2) and b) mean ρ of the province sheltering the reserve of interest. c) Mean number of interconnected MPAs (i.e. number of reserves situated within the same hydrodynamical province). Light grey contours represent the 200 m isobath. Results using a PLD of 45 and 60 days are qualitatively similar with a slight increase of the mean area and the number of interconnected reserves. Note that non-contiguous areas belonging administratively to a given reserve were treated here as a single MPA.

poral variability (not shown). MPAs implemented within narrow continental shelves bounded by energetic currents are characterized by rather large provinces ($\geq 7 \times 10^4\ km^2$) and moderate coherence ($0.65 \leq \rho \leq 0.8$). These reserves are situated along the French Côte d'Azur with the Liguro-Provencal Current, the Catalan coast with the Northern Current, the Moroccan/Algerian coastlines impacted by the Algerian Current and in the eastern Levantine basin with the jet-like intensifications of its gyre circulation (Millot and Taupier-Letage, 2005). This elevated connectivity is driven by the adjacent currents that regularly intrude onto the shelf, enhancing larval dispersal along the current axis, as suggested by the numerous interconnected MPAs (≥ 15) along the French, Catalan and Israelian coastlines. In contrast, MPAs situated within extended continental shelves, such as the Gulf of Lion and the Tunisian/Libyan shelf, are characterized by small dispersal area ($\leq 6 \times 10^4\ km^2$) and large coherence ($\rho \geq 0.8$). Unless exceptional intrusion events (not shown), the inner-shelf is isolated by shallow bathymetry holding the current off the shelf break, thus resulting in restricted connectivity. Note that most MPAs associated with narrow shelves and sluggish circulation (such as the Tyrrhenian, Corsican and Sardinian coastlines) behave quite similarly to the latter group with small dispersal surfaces.

5.5

Conclusions

To conclude, we found that despite the stochastic nature of larval dispersal (Siegel et al., 2008), the extraction of provinces organizing surface dispersion, allows to highlight local oceanographic characteristics resulting in the emergence of connectivity regimes. They should be in accord with the main conservation objectives to ensure successful implementations of coastal and offshore marine reserves. For instance, the allocation of MPAs within narrow shelves bounded by currents would favor larval export over large distances (Pelc et al., 2010) whereas reserves created within internal seas or large continental shelves would rather promote the restoration of local populations (Pineda et al., 2007). Overall, the Mediterranean MPAs are not evenly distributed across the spatial partitioning of the seascape revealed by our analysis. Moreover, the "size and spacing" guidelines already studied theoretically, may differ depending on the local dispersal behavior. Our results suggest the use of few large MPAs located in each stable hydrodynamical unit of the western Mediterranean basin and the Adriatic sea whereas numerous small MPAs evenly distributed across the fluctuating units might be preferable in the Ionian and Aegean seas.

CHAPTER 6

Local network measures and population dynamics

6.1

Characterizing metapopulations and subpopulations in marine seascapes

Marine populations are commonly structured as a ”metapopulation” in which discrete ”subpopulations” are linked to each other via the exchanges of individuals (adults, juveniles, larvae, eggs...) (Cowen and Sponaugle, 2009; Calò et al., 2013). This *population connectivity* (i.e. the exchanges of individuals among subpopulations) can be evaluated with the study of *genetic connectivity*, defined as the degree to which gene flow affects evolutionary processes within subpopulations. Another closely related concept is the *demographic connectivity* which measures the impact of dispersal processes on growth and survival rates of a subpopulation (Lowe and Allendorf, 2010). A good understanding of population connectivity, together with the genetic and demographic connectivities, are fundamental for predicting the replenishment of a given site after disturbance (population persistence), for the maintenance of genetic diversity within populations and overall for an efficient management of marine ecosystems (Palumbi, 2003; Cowen et al., 2006; Hastings and Botsford, 2006; Cowen and Sponaugle, 2009).

Depending on the rate of colonization, previous studies distinguished *Open* subpopulation that receives/exports individuals from/to other subpopulations while *Closed* subpopulation does not exchange individuals to an appreciable extent. This implies that open subpopulations are primarily maintained through network persistence (i.e the exchange of individuals with neighbouring subpopulations) while closed subpopulations mainly survive through self-persistence (i.e local birth rate higher than death rate). Another distinction among open subpopulations that are geographically isolated was introduced by (Pulliam, 1988) with the concept of *Source/Sink* dynamics. It arose from the spatial variability of habitat quality which leads to different demographic and exportation/importation rates in each subpopulation (Roberts, 1997; Crowder et al., 2000). (Pulliam, 1988) focused on terrestrial species with low dispersal so that the source/sink character could be estimated by local demographics. Despite the high dispersal ability of early life stages of marine species and the tremendous heterogeneity of oceanic habitats, (Cowen and Sponaugle, 2009) defined a source population as a subpopulation in which the net export of individuals is greater than the net import; the reverse is a sink.

Estimating the extent to which subpopulations are open/closed or source/sink has lead to define various local connectivity metrics that are directly or indirectly related to larval transport. A particular attention was given to Self-Recruitment (SR), which measures the proportion of all local larval recruits that originated from the source population (e.g. Planes et al., 2009; Saenz-Agudelo et al., 2011). Other studies aimed at estimating the Local Retention (LR): it is defined as the proportion of local larval production retained on a site (Carson et al., 2011; Hogan et al., 2012). Because SR is a function of the number of larvae arriving from elsewhere and LR depends on the number of those leaving, these two metrics do not inform population connectivity similarly. Concerning the proxies of exchange, (Crowder et al., 2000; Bode et al., 2006) showed the importance of appraising source/sink dynamics for conservation purposes. Some authors proposed to simplify the source/sink character of a given subpopulation by the directional movement of larvae in the currents since larvae have been identified as the most dispersive stage for many marine species. (Roberts, 1997) defined a source subpopulation when its larval export is greater than larval import, the reverse is a sink. Coupling hydrodynamical and population models, added the concept of population persistence by considering that a source must persist without larval supply, contrary to the sink which depends on it. Despite the crucial information brought by these open/closed and source/sink proxies applied to larval transport, they have not been studied yet. Furthermore, although they are often employed to describe the state of a local population or even to make management recommendations, the link with oceanographic processes and their inter-relationship remains unclear.

6.2

Defining local network measures as proxies of larval connectivity

Here we present a new modeling framework that allows characterizing larval dispersal and its impacts on the spatial structures and dynamics of marine populations from a large-scale perspective (Dubois et al., 2015). Although this framework is equally applicable to any target specie with well-known biological traits and to any oceanic areas, the connectivity patterns evidenced here are, under some assumptions, relevant for several key organisms of Mediterranean ecosystems. Following the method explained in Part I we are able to construct several networks from which various connectivity proxies are extracted. We retained two PLDs of 30 and 60 days and both winter and summer spawning by considering three successive starting times t_0 for each season ($t_0 = 1^{st}$, 15^{th} and 31^{st} of January and $t_0 = 1^{st}$, 15^{th} and 31^{st} of July) to account for episodic and variable spawning events. A total of 120 numerical experiments (6 starting times over 10 years for 2 PLDs) were performed to analyze robust basin-scale connectivity patterns focusing on species with a wide geographical range and potential for large-distance dispersal.

We aim here at describing the flow of larvae between all geographical sub-areas of the surface ocean (i.e. single nodes of the network), focusing on both the pathways and the fluxes. In our Network Theory perspective, this is equivalent to focus on centrality measures of single nodes. For each node, one can examine the number of links (*degree*) and the sum of relative weights (*strength*) emanating from it, referred to as the *OUT-degree* and *OUT-strength*, and those arriving in it, called the *IN-degree* and *IN-strength*. We define P_{ij} and L_{ij} as the weight and the adjacency matrices of the transport network under study.

Botsford et al. (2009) defined the Local Retention (LR) as the proportion of locally produced settlement to local larval release. Since the local larval release and the success of recruitment are here assumed to be homogeneous in space, LR can be approximated by the proportion of locally retained particles (i.e. the diagonal elements of the connectivity matrix):

$$LR_i = P_{i,i}\,. \tag{6.1}$$

Botsford et al. (2009) defined the Self-Recruitment (SR) as the ratio of locally produced settlement to settlement of all origins at a given site. With our assumptions, SR corresponds to the ratio of retained particles to the total incoming particles from all origins (including those produced locally):

$$SR_i = \frac{P_{i,i}}{\sum_{x=1,x\neq i}^{N}(P_{x,i}) + P_{i,i}} = \frac{P_{i,i}}{\sum_{x=1}^{N}(P_{x,i})}\,. \tag{6.2}$$

For each node, we also compute the local *OUT-strength* which measures the proportion of particles released locally that were transported elsewhere. It is calculated by summing the rows of the connectivity matrix, excluding the diagonal element. Similarly, the local *IN-strength* corresponds to the proportion of particles arriving in a given site originating from elsewhere. It is computed by summing-up the columns of the matrix, excluding the diagonal element.

$$OUT_i^s = \sum_{x=1,x\neq i}^{N}(P_{i,x}) = \sum_{x=1}^{N}(P_{i,x}) - P_{i,i}\,. \tag{6.3}$$

$$IN_i^s = \sum_{x=1,x\neq i}^{N}(P_{x,i}) = \sum_{x=1}^{N}(P_{x,i}) - P_{i,i}\,. \tag{6.4}$$

We also computed the *IN-degree* (*OUT-degree*, respectively) which represents the number of incoming (outgoing, respectively) links into (from, respectively) a given site:

$$OUT_i^d = \sum_{x=1,x\neq i}^{N}(L_{i,x}) = \sum_{x=1}^{N}(L_{i,x}) - L_{i,i}\,. \tag{6.5}$$

$$IN_i^d = \sum_{x=1,x\neq i}^{N}(L_{x,i}) = \sum_{x=1}^{N}(L_{x,i}) - L_{i,i}\,. \tag{6.6}$$

Then, we examine the relative importance of larval export versus import to inform the local source/sink dynamics. The *SourceSink* (SS) metrics are thus defined by:

$$SS_i^d = \frac{IN_i^d}{IN_i^d + OUT_i^d}\,. \tag{6.7}$$

$$SS_i^s = \frac{IN_i^s}{IN_i^s + OUT_i^s}. \tag{6.8}$$

By studying this ratio on both *strength* and *degree* variables, it evaluates the source/sink character both in terms of the total amount of larvae (*strength*) and in terms of the spatial diversity of origins/destinations (*degree*). Note that these metric are restricted between 0 and 1 by construction and its negative/positive deviation from 0.5 allows us to identify source/sink, respectively. A node characterized by $SS_i^s \ll 0.5$ can be considered as a net source of larvae whereas a node with $SS_i^s \gg 0.5$ represents a net sink. Similarly, a node characterized by $SS_i^d \ll 0.5$ indicates that it exports larvae to many different downstream nodes but it receives from just a few. A node with $SS_i^d \gg 0.5$ represents a site that receives larvae from many different upstream sites but that exports somehow unidirectionally to just a few locations.

Spatial patterns of each connectivity metric in the Mediterranean basin are studied through the mapping in geographical coordinates of their local (node-by-node) time-averages across all 120 simulations. Since a significant intra-annual variability of the oceanic circulation has been documented in the Mediterranean sea (e.g. Bakun and Agostini, 2001; Pinardi et al., 2013), we also investigated how these spatial patterns change depending on the season of interest. To do this, we computed temporal averages over a subset of the simulations considering the two spawning seasons separately, i.e. averaging the connectivity proxies using only winter or summer matrices.

To quantify the temporal variability of the local connectivity metrics, we measured the amount of variation from the temporal means μ by computing the standard deviation σ. We then relate locally the standard deviation to the mean to evaluate the confidence to be given at the averages reported on the maps. The local mean is considered non-significant if $\sigma > \mu$ (see the black crosses on Fig. 6.1 and 6.2). Those nodes are indeed characterized by a particularly high temporal variability, indicating oceanic regions where the stochasticity of the circulation prevent any robust characterization of connectivity.

The main conclusions of this analysis are reported hereafter:

- There are more than 50% of nodes that always show null LR/SR and less than 3% with LR/SR > 50%.
- For *IN/OUT-degree*, we find that on average 50% of nodes exchange larvae with more than 35 upstream and downstream sites; about 20% of boxes are highly connected with 70 or more upstream and downstream links.
- For *IN/OUT-strength* about 40% of all sites receive more particles than their initial quantity, thus behaving always ”larval sink”.
- The PLD is the most important factor of variability, especially for retention indices.
- Long PLDs increase the dispersal potential, as evidenced by stronger larval fluxes (*strength*) and more anisotropic transport with higher diversity of destinations/origins (*degree*).
- The seasonal and inter-annual variability of our proxies have similar magnitude.

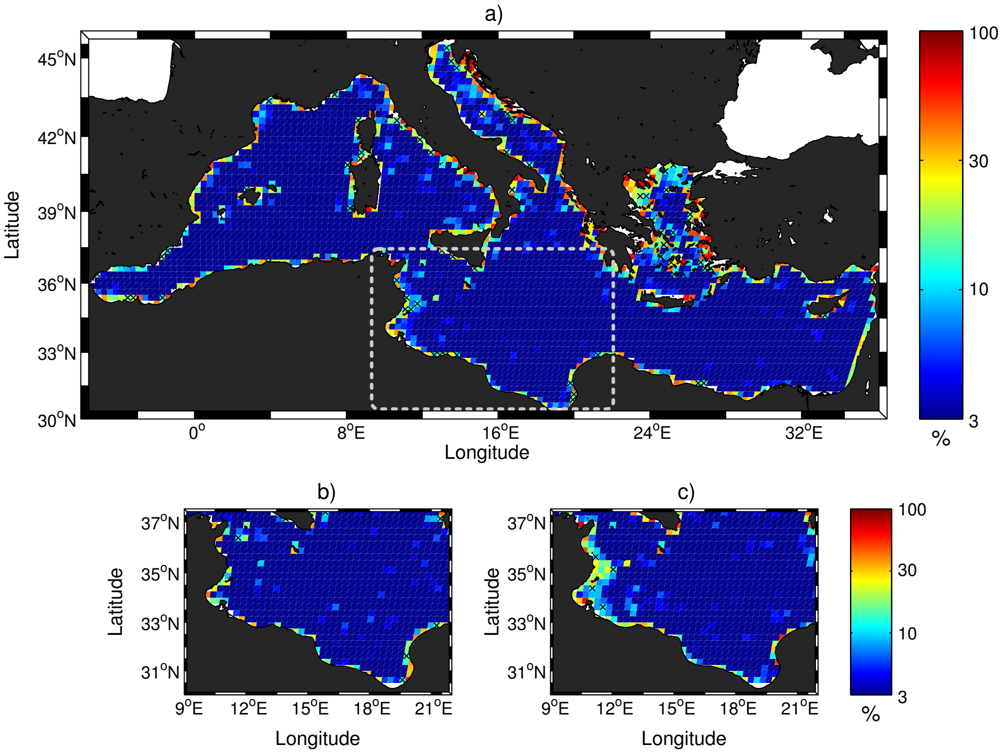

Figure 6.1: (a) Mean Local Retention (LR) for a PLD of 30 days averaged across 60 (winter and summer) connectivity matrices; (b) mean LR over the southern Ionian sea averaged across 30 winter matrices and (c) 30 summer matrices. Note that open ocean boxes are excluded from this variability analysis as their weak LR/SR (< 10%) are not analysed here.

Despite significant variability, maps of mean LR and SR over the Mediterranean basin (Fig. 6.1) reveal robust patterns of retention rates. The largest LR/SR (> 30%) are observed along the continental and insular coastlines while moderate values (10% < LR/SR < 30%) may be also find slightly offshore in boxes located over the continental shelf (depths < 200 m). In contrast, most open ocean boxes (depths > 200 m) show on average null or very low (> 1%) LR/SR (Fig. 6.1a).

While these global patterns are observed for all simulations, some noticeable seasonal differences can be highlighted. For instance, elevated values of LR and SR are found over the Tunisian shelf during summer whereas they are restricted to the near-coastal boxes in winter (Fig. 6.1b and c).

All σ computed over 60 (global mean; Fig. 6.2) or 30 (seasonal mean; Fig. 6.3) connectivity matrices are of the same order of magnitude for both PLDs and reveal that less than 1% of boxes have $\sigma > \mu$. Their locations do not show any consistent pattern, except in the western Alboran sea which has been disregarded from our analyses [1].

$SS - degree$ and $SS - strength$ show very identical spatial patterns with slightly smaller magnitude for for larval sinks in the $SS - strength$ variable (Fig. 6.2); this is true for both PLDs (not shown). As such, we only analyse the patterns observed in both $SS - degree$

---

[1] This methodological artefact is due to the lack of constant particles seeding despite the continuous entrance of Atlantic water through the Gibraltar strait.

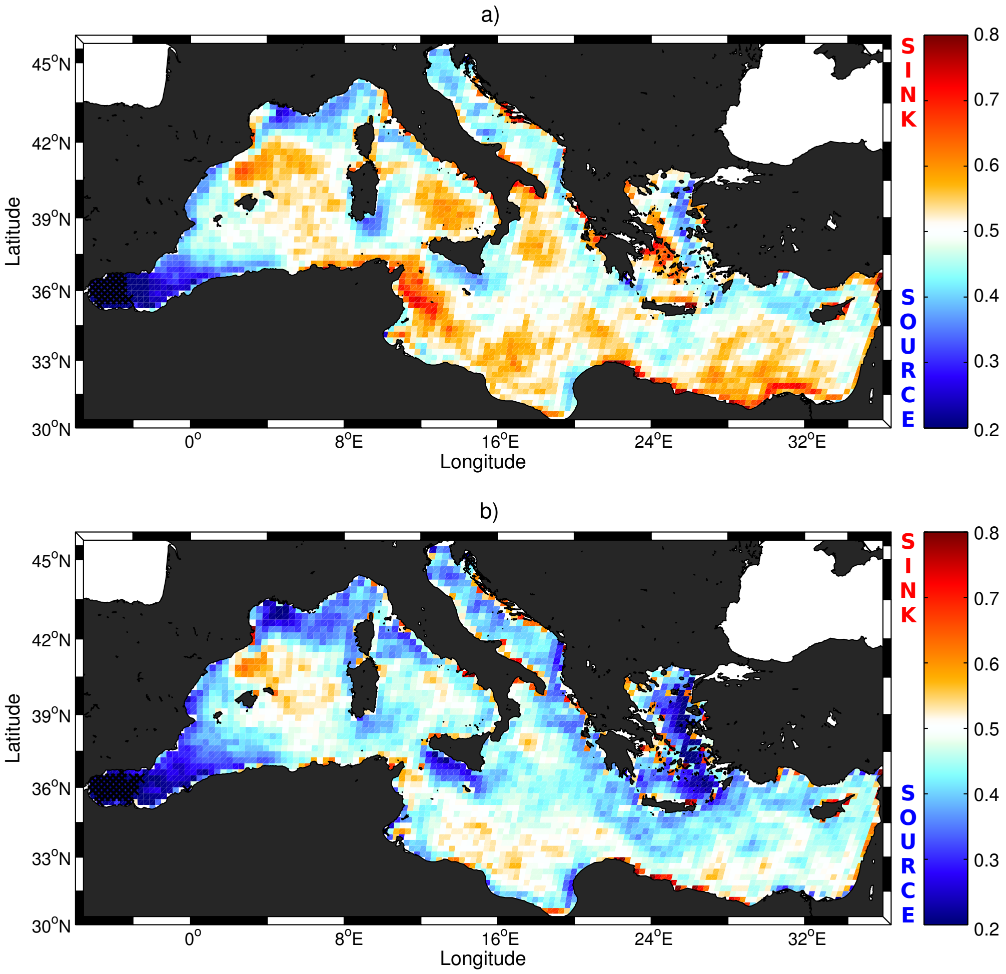

Figure 6.2: (a) Mean *SourceSink−degree* and (b) mean *SourceSink−strength* averaged across 60 (winter and summer) connectivity matrices for a PLD of 60 days.

and *SS−strength* using a PLD of 60 days (as longer dispersal tends to accentuate patterns of exchange).

Some regions are characterized by relatively stable behaviour throughout the year (Fig. 6.2). For instance, the Ligurian sea is mainly marked by a strong export of larvae (source) restricted to the near-shore areas (Côte d'Azur) in winter and extending offshore during summer (Fig. 6.3a, b). The Gulf of Lion can be divided into two subregions: the north-eastern areas which behave as a larval source with larval exportation higher than importation whereas the south-western coastlines are larval sinks. These patterns of *SS − degree*, as well as of *SS − strength* to a lesser extent, are quite similar in both seasons but the distinct behaviours are emphasized during winter (Fig. 6.3a, b). The Adriatic sea has a persistent larval sink along the southern Italian shores (Puglia coasts and the gulf of Taranto). Coastlines of Lybia, Egypt, Israel and Lebanon behave mainly as larval sinks throughout the year (Fig. 6.2), with an intensification along the Egyptian coastlines in summer (Fig. 6.3a, c).

Others parts of the Mediterranean sea are marked by two very distinct seasonal patterns. For instance, during summer most of the Balearic sea behaves as a large larval sink (Fig.

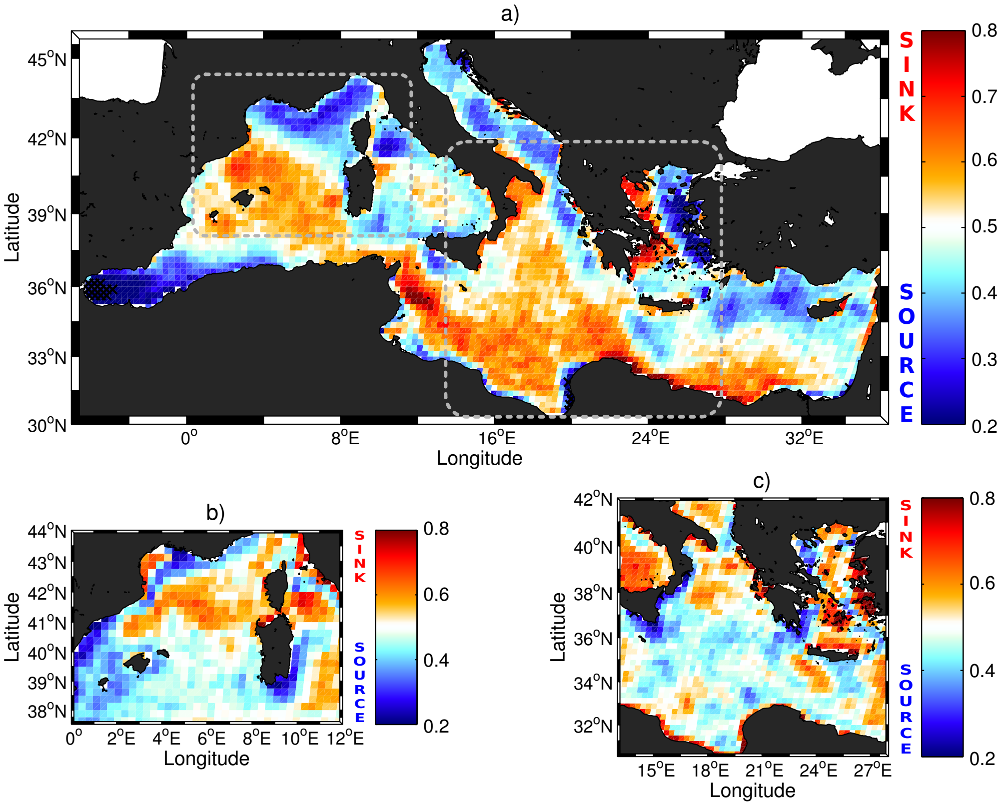

Figure 6.3: (a) Mean *SourceSink* − *degree* averaged across 30 summer matrices and mean *SourceSink*−*degree* averaged across 30 winter matrices (b) over the north-west Mediterranean and (c) over the Ionian sea for a PLD of 60 days.

6.3a, b), whereas the winter results in a different pattern with a pathway behaving as a clear larval source. The northern Ionian sea and the coastlines of Calabria and Sardinia behave during winter as a source while the western Greek shores are sinks (Fig. 6.3c). During summer, there is the opposite tendency with larval sources observed off Greece (westward). The same pattern is observed in the Aegean sea (Fig. 6.3a, c): in winter its eastern side acts as a larval sink while its western side is a source. The reverse is observed in summer. At smaller scales, the Gulf of Taranto in wintertime acts as a sink on its north-eastern coast and as a source on its south-western side but it appears as a strong sink in summer. Another clear seasonality exists in the coast of Gulf of Sirte which behave as sink in winter and source in summer (Fig. 6.3a, c).

Finally the central Ionian sea, central Levantine sea and Algerian basin are characterized by heterogeneous patterns of both *SS* − *degree* and *SS* − *strength* (Fig. 6.2), with particularly weak signal close to 0.5, making it difficult for us to depict clear source or sink in these open ocean regions.

Investigating the relationship between LR and SR is important because it is generally easier to estimate empirically SR (which only depends on local estimates) rather than LR (which depends on the highly uncertain larval production of both local and remote subpopulations). When plotting LR versus SR when considering each box and each simulation independently, we find a rather loose correlation (Fig. 6.4a) with a R^2 of 0.69 indicating a significant spread of the cloud of points. However, by averaging locally the

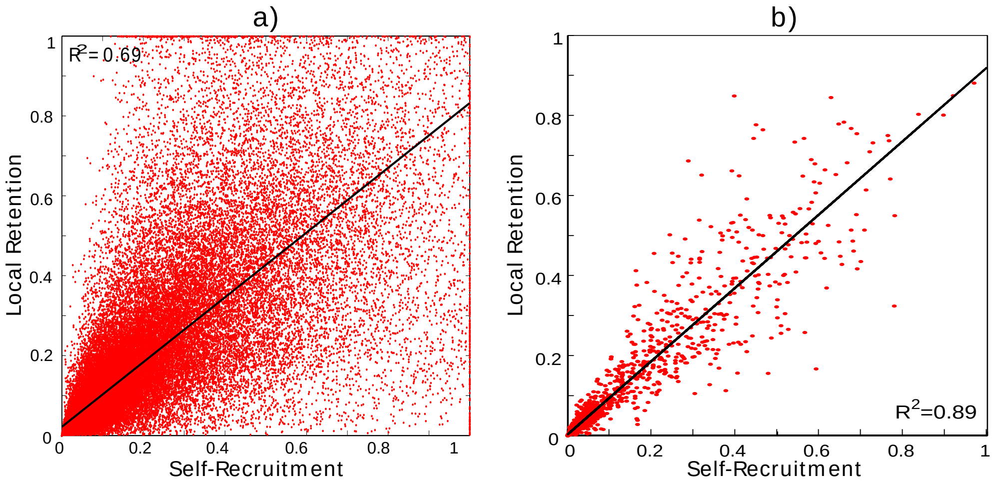

Figure 6.4: Relationship between LR and SR for a PLD of 30 days when (a) considering each box (3270) and each simulation (60) independently, and (b) when averaging our proxies locally across all simulations.

LR and SR values over all simulations (for a fixed PLD), LR and SR appear positively correlated through a more compact linear relationship (Fig. 6.4b). In this case, the R^2 coefficient is 0.89, suggesting that LR and SR are in average almost equal at a given site (under our assumptions of homogeneous larval production and mortality).

To investigate if there is a global relationship between retention and exchange metrics we distinguished each point of the LR/SR scatterplots according to its value of $SS-degree$ and $SS-strength$ (Fig. 6.5). We find that a sink site (receiving more larvae than sending them) tends to have higher LR than SR whereas a source site is characterized by SR higher than LR. A tight linear relationship between SR and LR (R^2 of 0.96) is found for the sites characterized by neutral SS (i.e. receiving as much larvae as sending them).

The basin-scale angle of our study allows highlighting common oceanographic conditions which would favour retention processes in any other oceanic region. Extended continental shelves and complex bathymetry (e.g. islands, capes, large bays) in association with relatively weak coastal currents lead to high retention near the coast. Exemplary locations of enhanced LR and SR are the Gulf of Lion, the Gulf of Valencia/Ebro delta, the Tuscan archipelago, the Croatian coastlines and the insular and near-shore subregions of the Aegean sea. This is consistent with the elevated retention processes that was documented in coral reef environments by small-scale bio-physical models (Treml et al., 2012). It is due to a sluggish mean circulation associated with (sub)mesoscale physical processes (e.g. topographically steered currents, trapped-eddies) whose characteristics time-scales are smaller or comparable with the duration of larval drift, thus limiting advection away from the shore. Retention is however diminished by large-scale energetic currents flowing close to the shore (e.g. the narrow continental shelves of the french "Côte d'Azur") and by long-lived mesoscale structures travelling offshore (e.g. the Algerian coast). The seasonal change of LR/SR observed over the Tunisian shelf illustrate well these two behaviours (Fig. 6.1b, c). The elevated LR observed during summer decrease in winter due to the acceleration and intrusion over the shelf of the Sicily-Strait-Tunisian-Current (Millot and Taupier-Letage, 2005).

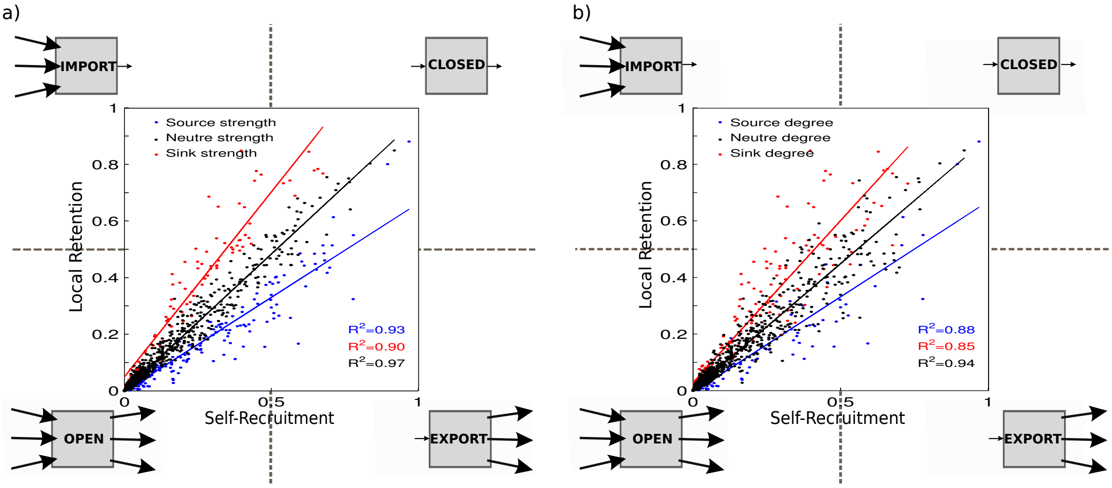

Figure 6.5: Inter-relationship among connectivity metrics for a PLD of 30 days. Mean relationship between LR and SR distinguishing ”source” (blue), ”sink” (red) and ”neutral” (black) box based on (a) $SS-strength$ and (b) $SS-degree$. Annotations represent schematically the behaviour of a node according to its location on the plot. ”Neutral” nodes have $SS-strength$ or $SS-degree \in [0.4, 0.6]$.

The variability of $SS-degree$ and $SS-strength$ metrics appears also primarily determined by the oceanographic setting as most larval sources/sinks identified here match well divergence/convergence zones (respectively). For instance, the ”sink” character of the Balearic sea could be related to the dominant wind regimes that forces convergent surface transport and downwelling in the centre of the gyre, especially in summer (Bakun and Agostini, 2001). In contrast, atmospheric patterns change during winter with the intensification of the Cierzo wind in the Ebro valley, resulting in a cyclonic wind stress curl producing an area of Ekman divergence and associated upwelling (Bakun and Agostini, 2001) in the southern part of the Balearic sea. The permanent retentive character of the north-eastern Balearic sea is consistent with who documented the highest residence times of surface drifters over 1992-2010 in this area. Other offshore sources, such as in the Adriatic and Tyrrhenian sea, are also reasonably well explained by the oceanographic context with the three cyclonic gyres in the Adriatic sea and the summer strengthening of the northern Tyrrhenian cyclonic gyre.

The impact of the pathway and state of a boundary current on connectivity properties is another relevant mechanism that is clearly evidenced in the north-western Mediterranean. The dominant circulation pattern in the region is the Liguro-Provencal Current (Ligurian sea) which then prolong into the Northern Current off the Gulf of Lion. This current intensifies during winter with a narrowing and strengthening of its transport (Millot and Taupier-Letage, 2005) that would favour larval export along its main axis. This explains well the narrow larval source observed in winter along the French coastline as well as the blue vein (source) further west that sharply separates the Gulf of Lion from the Balearic sea. In contrast during summer, the source region (Ligurian sector) extends further offshore due to the slower and less defined boundary current, in association with the ”Mistral” wind which tends to veer along the coast forcing coastal upwelling (Bakun and Agostini, 2001).

The Gulf of Lion is characterized by a shallow bathymetry that often maintains the core of the Northern Current off the shelf-break. As such, the inner-shelf circulation is primarily influenced by the ”Tramontane” and ”Mistral” winds which produce a divergent zone associated with upwelling along the Provence/Camargue coastlines (north-eastern parts of the Gulf) and a convergent zone with downwelling along the Languedoc-Roussillon coast (south-western parts of the Gulf). These divergence/convergence zones match very well the source/sink dynamics described previously. Note that the accentuated source/sink pattern observed in the Gulf during winter is also consistent with the seasonal intensification of ”Mistral” and ”Tramontane” winds. Similarly, other ”dipoles” of upwelling/downwelling forced by seasonal wind regimes are in accord with our proxies (Fig. 6.2, 6.3). The source/sink dipole documented along the northern Ionian and Aegean coastlines are both forced by summer southward Etesian winds (Bakun and Agostini, 2001); the winter dipole observed in the Gulf of Taranto is due to the winter southward Bora winds. The larval sinks observed along the Lybian/Egyptian shores are due to the dominance of coastal convergence all year long and the seasonal sources observed along the coastlines of the Gulf of Gabes and Sirte are related to the spring/summer wind reversals (Bakun and Agostini, 2001).

Although relatively stable circulation can be consistently associated with larval sources/sinks, it is worth mentioning that some oceanic regions such as the central Levantine sea and the Algerian basin are characterized by chaotic and turbulent transport (Millot and Taupier-Letage, 2005) preventing a clear depiction of their source or sink behavior.

6.3

Hydrodynamical and genetic connectivity for management purposes

There is growing evidence that oceanographically-induced barriers or preferential conduits of larval transport play a key role in shaping the genetic structures of marine organisms of different taxa (e.g. Banks et al., 2007; White et al., 2010; Mokhtar-Jamaï et al., 2011; Schunter et al., 2011; Franchini et al., 2012; Soria et al., 2012; Teacher et al., 2013). Nevertheless, no consensus has been yet reached about whether the observed genetic structure reflects primarily gene flow patterns (supposedly mediated by larval transport), genetic drift or local adaptation through mutations. We suggest that our results, and more generally the tunable modeling framework we proposed here, offer great opportunities to help addressing this long-standing controversy.

In particular, geneticists have been interpreting their datasets with the concepts of ”open/closed” and ”source/sink” populations but there is no universal agreement on the genetic structures supporting these behaviors. Our basin-scale maps of larval retention and exchange provide quantitative information to formulate and test improved hypothesis about the dominant gene flow patterns and the subsequent categorization of a subpopulation as ”open/closed” and ”source/sink”. Another important shortcoming of genetic studies is the discrete sampling strategies. Sampling of individuals is often opportunistic (i.e. where samples are available) or is designed to test a supposed barrier to gene flow (e.g. the Oran-Almeria front). The maps presented here could be used to plan intelligently the sampling of population genetic studies by targeting sub-regions with dissimilar connectivity behaviors (as predicted by our model) which could be then evaluated based on the genetic dataset.

The persistent larval sources/sinks that we identified have the potential to impact the genetic connectivity of local subpopulations. For instance, the genetic homogeneity found among the sea-bream *Diplodus sargus* populations of the Italian, French and Spanish coasts (Lenfant and Planes, 1996) could be related to the efficient gene flows occurring among those source areas. In contrast, genetic samples originating from the south-western part of the Gulf of Lion were found differentiated from the surrounding areas for *D. sargus* (Lenfant and Planes, 1996) and the sessil specie *Paramuricea clavata*. These authors hypothesized an oceanographically-induced gene flow to explain the genetic differentiation observed in the south-western Gulf of Lion. It is confirmed by a small-scale study (Guizien et al., 2012) and by our basin-scale model, both predicting a marked larval sink in this region.

Note however that with our assumptions, which differ from those of Pulliam (1988), a subpopulation is characterized as a sink under two contrasting situations. On one hand, a subpopulation is classified as a sink when it sends almost no emigrant elsewhere (i.e. weak exportation, high LR), suggesting it is genetically influenced by itself. In this case, it is the genetic drift due to local adaptation that will result in a genetic differentiation with its neighbouring subpopulations Hartl and Clark (1998); Jolly et al. (2009). On the other hand, a subpopulation is a sink when importation from neighbouring source(s) are particularly high (i.e. low SR), indicating a significant input of genetic materials from surrounding subpopulations and possibly resulting in a weak genetic differentiation between the connected source/sink subpopulations Gaggiotti (1996). Such sink population can be further distinguished into those influenced by a single source or those who received from several sources, potentially maximizing its genetic diversity (Pannell and Charlesworth, 1999). Indeed subpopulations of the Balearic sea have been often described as a "mixed genes pool" with diverse genetic influences from the Alboran, Catalan and Ligurian seas (Rozenfeld et al., 2008; Ledoux et al., 2010; Mokhtar-Jamaï et al., 2011; Hamdi et al., 2012). Our results consistently suggest that the Balearic sea is a basin of attraction for drifting larvae as evidenced by a strong influx (*strenght*) from diverse origins (*degree*), potentially resulting in a mix between local genes and those from the surrounding sources. Furthermore, Chaoui et al. (2009) documented a clear subdivision between the French and Algerian coastlines for the highly dispersive specie *Sparus aurata*. In that case, both the permanent larval sink in the north-western Balearic sea and the presence of the Balearic front extending eastward (Rossi et al., 2014) could act as efficient basin-scale barriers to gene flow.

Discussing further these issues is not the scope of this chapter. However it clearly highlights the importance to analyze together these complementary connectivity metrics at large-scale, as is allowed by our method, to properly characterize marine subpopulations.

As recently suggested by (Lett et al., 2015) (using a metapopulation model and the genetic dataset of (Saenz-Agudelo et al., 2011)) and the population genetic study of (Hogan et al., 2012), our results confirm that SR may be a good predictor of LR especially when averaging the stochasticity of the oceanic circulation at each site. Considering that previous similar conclusions were based on genetic dataset, it suggests that the variability of larval dispersal does not impact strongly the genetic structure of marine metapopulation but it is rather the mean dispersal patterns that dominate the genetic imprint.

Furthermore, our results further advocate that this compact linear relationship holds when the local source-sink dynamic is relatively ″neutral″, i.e. when a site is not identified as a strong larval source or sink. In this case LR, which can help assessing population persistence (Burgess et al., 2014), can be well approximated by SR (generally easier to measure by empirical approaches). In contrast, when a population is characterized by large relative differences between its exportation and importation, LR and SR are still linearly correlated but with a coefficient > 1 for the sinks and < 1 for the sources. This is because a site characterized by high SR (i.e. low ″Openness″ (Hixon et al., 2002)) rarely receives any larvae from elsewhere (importation) and inversely, a site with high LR would have very low exportation. Consequently, a subpopulation with high SR and LR can be considered as ″closed″ due to the limited upstream and downstream exchanges it has with distant sites or subpopulations (see annotations on Fig. 6.5). Similarly, ″open″ subpopulation would be characterized by low SR and LR suggesting high import and export, respectively.

In other words, the greater is the relative difference between SR and LR, the larger are the deviations of $SS-degree$ or $SS-strength$ from 0.5, more pronounced is the source or sink behavior. Indeed, almost all intermediate behaviors were observed in our simulations, especially in the non-averaged case (Fig. 6.4.a). For instance, a site presenting simultaneously $SR < 0.5$ and $LR > 0.5$ (upper left corner of the plots in Fig. 6.5a) will systematically import more particle than it export and will thus behave as a sink. The inverse reasoning stands for the source sites. While these conclusions based on $SS-strength$ are straightforward, we show that they also hold when considering $SS-degree$: a site which tends to import from distinct geographical regions but export to only a few (a sink node in terms of $degree$) is also characterized by LR higher than SR, having possibly consequences for the local genetic and phenotypic diversity.

The integrated interpretation of retention and exchange large-scale connectivity indices has implication for the implementation of coastal and pelagic Marine Protected Areas (MPAs). They all inform how subpopulations are connected through larval transport so that they would help predicting the effects of protection measures or of external perturbation on both local and surrounding subpopulations. For instance, the protection of a site whose larval supply originates from several non-protected ″upstream″ subpopulations may be inadequate. To assure the persistence of marine populations, a fundamental objective of ecosystems conservation and fisheries management (Burgess et al., 2014), MPAs could be located targeting source (divergent) areas (Roberts, 1997; Crowder et al., 2000) with moderate retention, thus favoring both self- and network persistence. Besides the identification of low threats - high diversity areas (Coll et al., 2015) or favorable habitats (Anadón et al., 2013), depicting both local and broad-scale connectivity of any oceanic area, as is allowed by our model, is a crucial pre-requisite for implementing future MPAs.

Note that rather than developing a highly realistic model that would apply to a single species, we selected here a range of parameters according to well-known ecological traits of various Mediterranean marine organisms to propose a new model which allows describing large-scale connectivity. The ″true″ connectivity of marine population depends however on numerous species-specific biological processes all of which are also influenced by abiotic factors such as the variable oceanic circulation and the heterogeneity of littoral and pelagic habitats (Cowen and Sponaugle, 2009). Despite resolving only some of these processes, we highlight specific oceanographic processes that control larval connectivity and that would affect both demographic and genetic connectivities of many marine organisms. Furthermore, it is worth emphasizing that the novel modeling framework presented here is indeed adaptable to any species whose biological traits and ecological preferences are precisely known.

Last, there exist numerical limitations associated with the resolution of the hydrodynamical model and of our transport network. Although near-shore areas are not particularly well resolved by the basin-scale model we used, some patterns are found in good agreement with regional studies based on dedicated high-resolution ocean model (Guizien et al., 2012).

6.4

Conclusions

To conclude, we stress that describing the connectivity of marine subpopulations, thought to be governed largely by the pelagic larval stage, has crucial ecological and managerial implications. We proposeed here a new modeling framework including Lagrangian simulations, tools from Network Theory and concepts of population dynamics, to better characterize marine connectivity at the scale of the Mediterranean basin, allowing to bridge the gaps between the distinct approaches commonly used.

Part III

OPTIMAL PATHS IN TEMPORAL FLOW NETWORKS

CHAPTER 7

Characterizing paths in temporal weighted networks

7.1

Posing the problem: transport and probability

Let's consider a temporal weighted directed network where weights are representative of transport of some quantity Q from one node to another. We suppose now that we do not have any constraint on the conservation of this quantity to treat the most general case. We define out-strength and in-strength of the i-node during a time interval $m = [t_{m-1}; t_m]$ as:

$$s^{(m)}_{out_i} = \sum_j \mathbf{A}^{(m)}_{ij} \,, \tag{7.1}$$

$$s^{(m)}_{in_i} = \sum_j \mathbf{A}^{(m)}_{ji} \,. \tag{7.2}$$

We consider the case of a single step, looking at the transport from the node I to the node J for $m = 1$. The amount of transported quantity Q will be represented by the weight $\mathbf{A}^{(1)}_{IJ}$ To simplify the notation we take the discrete times to be integer: $\{t_0, t_1, ...\} = \{0, 1, ...\}$. We can now pose two different fundamental questions:

1. What is the fraction of Q leaving I at time $t = 0$ that end in J at time $t = 1$ (taking as reference the total quantity of Q present in I). Or equivalently, what is the probability $P(J_{t=1}|I_{t=0})$ for a random walker to end in J at time $t = 1$ under the condition of starting from I at time $t = 0$?

2. What is the fraction of Q present in J at time $t = 1$ that come from I at time $t = 0$ (taking as reference the total quantity of Q present in J). Or equivalently, what is the probability $P(I_{t=0}|J_{t=1})$ for a random walker to come from I at time $t = 0$ under the condition of being in J at time $t = 1$?

To address to these questions we need the information carried by $s^{(m)}_{out_i}$ and $s^{(m)}_{in_i}$ to normalize the weight properly.

1. We divide the weight by the total amount of Q exported from I during $m = 1$:
$$p_{IJ}^{(1)} = P(J_{t=1}|I_{t=0}) = \frac{\mathbf{A}_{IJ}^{(1)}}{s_{out_I}^{(1)}} \,. \tag{7.3}$$
2. We divide the weight by the total amount of Q imported to J during $m = 1$:
$$f_{IJ}^{(1)} = P(I_{t=0}|J_{t=1}) = \frac{\mathbf{A}_{IJ}^{(1)}}{s_{in_J}^{(1)}} \,. \tag{7.4}$$

We stress that the two quantities are related by time-reversal transformation. If the system is invariant under time-reversal we find indeed $P(J_{t=1}|I_{t=0}) = P(I_{t=0}|J_{t=1})$ and so $p_{IJ}^{(1)} = f_{IJ}^{(1)}$.

To unveil the time-dependent connectivity features of our temporal weighted network, we need some tools to describe transport dynamics across several temporal snapshots of the network.

First, we introduce the concept of *extremal path* η_{IJ}^{M}, as the path of M-temporal steps connecting the pair of nodes (I, J) maximizing some quantity relevant for the process in study. If the classical concept of shortest path is related to the idea of minimizing the distance connecting two nodes, here we are not anymore focused on a geometric interpretation. We are looking instead at the most efficient way to connect two nodes independently on the length of that connection.

Until this point we did not need any assumptions about the memory of the transport processes across the network, but from now we will assume that such processes are Markovian. Under this assumption and considering that both probabilities and fractions are multiplicative quantities we could extend to an arbitrary number of steps M the definitions for $p_{IJ}^{(1)}$ and $f_{IJ}^{(1)}$. We note that probabilities and fractions are multiplicative when:

1. Probabilities of paths are referred to the probability for a random walker moving on the network of going across a specific path connecting two nodes. We assume that the random walker is moving according the flow of Q but it represents an independent subject and it could not be seen as a parcel of Q. In particular it is not involved in the creation/destruction dynamics.
2. Fractions of exported Q refer to the fraction of the quantity Q belonging to a fixed starting node that is present in a destination node after going across a specific path. Here we assume that across each time step in every node the fractions of Q are conserved i.e. creation/destruction dynamic conserves the proportions of Q originated in different places.

Probabilities and fractions are the two quantities that will be maximized to identify *extremal path* in our network.

Considering the time interval $[t_0, t_M]$ divided into M-steps we define an *extremal paths* as the $(M + 1)$-tuple of nodes among the set of all possible paths $\mu = \left\{\langle I, k_1, \ldots, k_{M-1}, J\rangle\right\}$

connecting those nodes, that maximize respectively:

$$(p^M_{I,J})_\mu = \frac{\mathbf{A}^{(1)}_{I,k_1}}{s^{(1)}_{out_I}} \left[\prod_{m=2}^{M-1} \frac{\mathbf{A}^{(m)}_{k_{m-1},k_m}}{s^{(m)}_{out_{k_{m-1}}}} \right] \frac{\mathbf{A}^{(M)}_{k_{M-1},J}}{s^{(M)}_{out_{k_{M-1}}}}\,. \tag{7.5}$$

Or:

$$(f^M_{I,J})_\mu = \frac{\mathbf{A}^{(1)}_{I,k_1}}{s^{(1)}_{in_{k_1}}} \left[\prod_{m=2}^{M-1} \frac{\mathbf{A}^{(m)}_{k_{m-1},k_m}}{s^{(m)}_{in_{k_m}}} \right] \frac{\mathbf{A}^{(M)}_{k_{M-1},J}}{s^{(M)}_{in_J}}\,. \tag{7.6}$$

So the *extremal path* η^M_{IJ} will be characterized by $M+1$ nodes in the network and the associated probabilities/fractions $P^M_{I,J} = \max_\mu\{(p^M_{I,J})_\mu\}$ or $F^M_{I,J} = \max_\mu\{(f^M_{I,J})_\mu\}$.

We also note that we can reconsider the concept of *extremal path* in the simpler case of static networks. In this case, indeed, even if we lose the temporal meaning of each step, the probabilistic interpretation remain valid.

Now we study how symmetries in the network allow us to find new relations between the two definitions of paths introduced above. Let's consider the case of local absence of sources and sinks:

$$s^{(l)}_{out_i} = s^{(l-1)}_{in_i} \quad , \quad \forall\ i, l\,.$$

It is easy to see that using this constraint we can express $p^M_{I,J}$ in function of $f^M_{I,J}$ and vice versa:

$$p^M_{I,J} = f^M_{I,J} \left(\frac{s^{(M)}_{in_J}}{s^{(1)}_{in_I}} \right), \tag{7.7}$$

$$f^M_{I,J} = p^M_{I,J} \left(\frac{s^{(1)}_{out_I}}{s^{(M)}_{out_J}} \right). \tag{7.8}$$

This means that $p^M_{I,J}$ and $f^M_{I,J}$ are proportional and the constant of proportionality only depends on properties of I and J and not on the path connecting them. The path among I and J that maximizes the two quantities will coincide.

In the case of double-stochasticity we have:

$$s^{(l)}_{out_i} = s^{(m)}_{out_j} = s^{(p)}_{in_k} = s^{(q)}_{in_z} \quad , \quad \forall\ i, j, k, z, l, m, p, q\,.$$

The relations become an identity:

$$p^M_{I,J} = f^M_{I,J}\,. \tag{7.9}$$

So in constant double-stochastic network the two concepts of probability/fraction coincide giving a connection with the concept of time-reversal invariance.

7.2

Most Probable Paths (MPPs) calculation

Here we will focus on the extremal paths maximizing the probability i.e. *most probable paths* among two nodes (MPPs) (Ser-Giacomi et al., 2015b,c). Let's write the probability

$(p_{I,J}^{M})_{\mu}$ as:

$$(p_{IJ}^{M})_{\mu} = \mathbf{T}_{Ik_1}^{(1)}\left[\prod_{l=2}^{M-1}\mathbf{T}_{k_{l-1}k_l}^{(l)}\right]\mathbf{T}_{k_{M-1}J}^{(M)}\,, \tag{7.10}$$

where:

$$\mathbf{T}_{k_{l-1}k_l}^{(l)} = \frac{\mathbf{A}_{k_{l-1}k_l}^{(l)}}{s_{out_{k_{l-1}}}^{(l)}}\,, \tag{7.11}$$

is the single step probability.

To find the MPP and its probability we use an adaptation of the Dijkstra algorithm (Dijkstra, 1959) which takes into account the layered and directed structure of our time-ordered flow graph. In brief, we realize that for on-step paths the maximum probability is $P_{Ik_1}^{1} = \mathbf{T}_{Ik_1}^{(1)}$ and then we apply recursively the formula:

$$P_{Ik_{l+1}}^{l+1} = \max_{k_l}\left(P_{Ik_l}^{l}\mathbf{T}_{k_lk_{l+1}}^{(l+1)}\right)\,, \tag{7.12}$$

for $l = 1, 2, M-1$ until reaching the endpoint $J = k_M$ (see Fig. 7.1 for a detailed explanation of the algorithm).

The cost of the computation is reduced by computing the accessibility (Lentz et al., 2013) of each intermediate node at each intermediate time. Once we fix initial and final nodes I and J we calculate at each intermediate time which are the nodes that are crossed at least by one path during the dynamics. A given node k_l at time t_l is crossed if it is reachable from I in l time steps, and at the same time, if from this node it is possible to reach node J in $M-l$ time steps. This information is collected in the so-called *accessibility matrices* (Lentz et al., 2013), which are properly time-ordered products of the network adjacency matrices **A**. Specifically, node k is contributing at time l to a path between I and J if and only if:

$$\left[\prod_{i=1}^{l}\mathbf{A}^{(i)}\right]_{Ik} \neq 0 \qquad \textit{and} \qquad \left[\prod_{i=l+1}^{M}\mathbf{A}^{(i)}\right]_{kJ} \neq 0\,. \tag{7.13}$$

Raising the number M of steps we observe a fast increase in the number of paths connecting two given nodes. It is thus crucial to understand how much the MPP is representative of the large set of possible paths joining two nodes. To assess in a quantitative way this issue we introduce the following quantity:

$$\lambda_{IJ}^{M} = \frac{P_{IJ}^{M}}{\sum_{\mu}(p_{IJ}^{M})_{\mu}}\,, \tag{7.14}$$

which determines the fraction of probability carried by the MPP with respect to the sum of probabilities of all paths connecting nodes I and J. Note that the denominator can be simply computed as the matrix-product entry $\left(\prod_{l=1}^{M}\mathbf{T}^{(l)}\right)_{IJ}$.

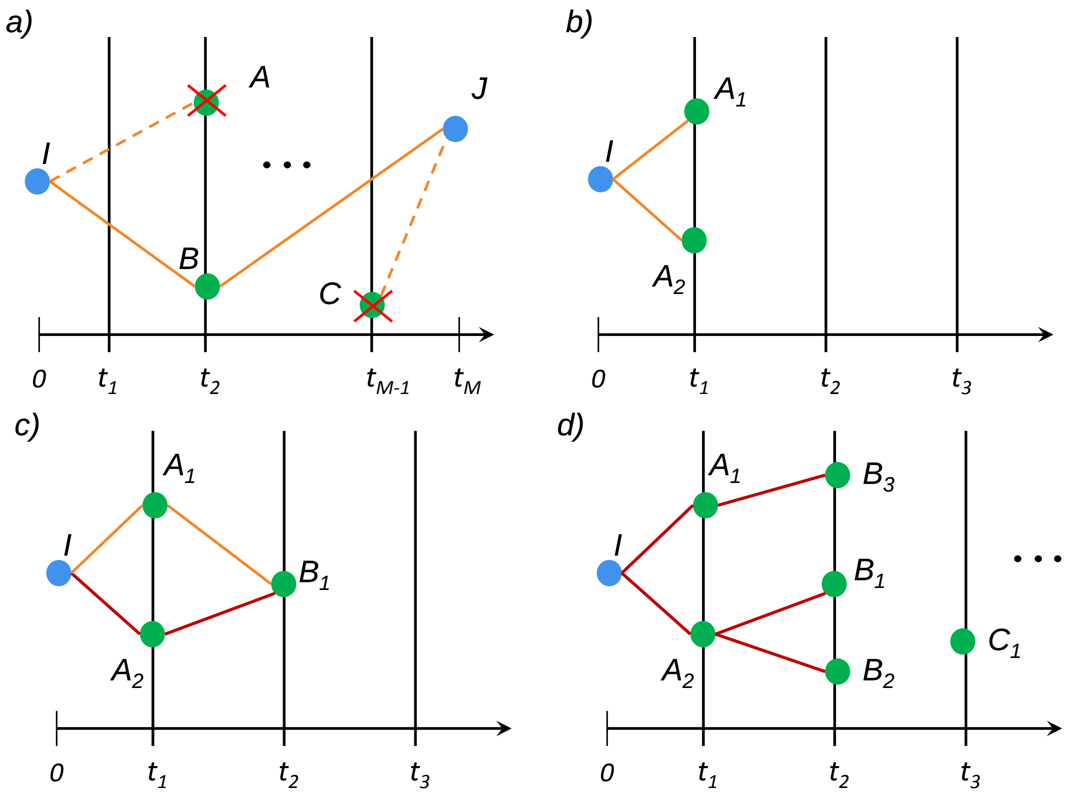

Figure 7.1: Schematics of the algorithm to find the MPP of M steps between I and J. a) First part: determination of the accessible nodes. Point A is reachable from I at $t = t_2$ but it is not possible to reach J from it in the rest of the time interval. Point C is not reachable from I at $t = t_{M-1}$ even if J can be reached from it. Point B satisfies both accessibility conditions, therefore, in contrast to points A and C, it belongs to the accessibility set and it will be considered in the calculation of the MPP. Systematic identification of all accessible nodes is done by applying the criteria in Eq. (7.13). The rest of the figure illustrates the recursive maximization procedure given by Eq. (7.12): b) In the first time step one assigns to the links towards the nodes A_1 and A_2 (considered to be the only ones in the accessibility set $\mathbf{U}_{IJ}^{(1)}$) the probabilities $\mathbf{T^{(1)}}_{I,A_1}$ and $\mathbf{T^{(1)}}_{I,A_2}$, respectively. c) For node B_1 one considers the links from A_1 and A_2, evaluates the path's probabilities $\mathbf{T^{(1)}}_{I,A_1}\mathbf{T^{(2)}}_{A_1,B_1}$ and $\mathbf{T^{(1)}}_{I,A_2}\mathbf{T^{(2)}}_{A_2,B_1}$, and selects the maximum one (in the figure the corresponding to the path I, A_2, B_1, red lines). One repeats this for all nodes B_1, B_2, B_3 in the accessibility set $\mathbf{U}_{IJ}^{(2)}$ to obtain the MPPs between I and these nodes, and then the procedure can be iterated again for the accessible nodes at time t_3.

7.3

Sets of highly probable paths

For large values of M, the MPP progressively loses dominance and, on average, does not carry a significantly high fraction of probability. However the dynamics, characterized by a high number of paths connecting initial and final points, can be still described by a few of them, which together have a non-negligible probability. To see this we can relax the definition of MPP and define a family of subsets of highly probable paths (HPP) holding most of the probability. In our formulation each subset $\mathcal{K}_{IJ}^{M}(r,\epsilon)$ is characterized by a *rank* $0 \leq r \leq M-1$ and a threshold parameter $0 \leq \epsilon \leq 1$. Ideally the sets would contain all the paths whose probability is larger than $\epsilon P_{I,J}^{M}$. But since exhaustive searching of all such paths becomes computationally prohibitive except for very small M, the second parameter r is introduced to determine the number of constraints imposed in the search for these relevant paths. Given the initial (I) and final (J) points we fix r nodes at intermediate times and look for paths between I and J made of segments which are MPPs connecting these intermediate nodes, by using the algorithm above. Different locations and times for these r intermediate nodes are scanned and paths with probability larger than $\epsilon P_{I,J}^{M}$ are retained and incorporated into the set $\mathcal{K}_{IJ}^{M}(r,\epsilon)$. For $\epsilon \to 1$, independently on the rank (or for $r = 0$) only the MPP is retained. $\mathcal{K}_{IJ}^{M}(r = M-1,\epsilon)$ contains all the paths with probability larger than $\epsilon P_{I,J}^{M}$. However, evaluation of these sets of HPPs can be computationally costly for high values of r, since the algorithm scales exponentially with r. Nevertheless interesting results can be obtained considering already low-order HPPs, i.e. $r = 1$ and $r = 2$.

Once one of the subsets is computed we can establish its significance by defining an extension of expression Eq. (7.14):

$$\lambda_{IJ}^{M}(r,\epsilon) = \frac{\sum_{\nu}(p_{IJ}^{M})_{\nu}}{\sum_{\mu}(p_{IJ}^{M})_{\mu}}, \tag{7.15}$$

where the sum in the numerator is over all the paths in the subset $\mathcal{K}_{IJ}^{M}(r,\epsilon)$ and the one in the denominator is over all paths connecting I to J.

When we deal with spatially-embedded networks (and this is the case of the applications that we will show later), another important aspect of the sets of HPPs is to establish how close, spatially, are they with respect to the corresponding MPP. This is obtained with an average distance function. Given two generic paths between initial and final points I and J, $\mu_1 = \{I, k_1 ..., J\}$ and $\mu_2 = \{I, l_1 ..., J\}$ we define their average distance as

$$\mathbf{d}(\mu_1,\mu_2) = \frac{1}{M-1}\sum_{i=1}^{M-1} d(k_i,l_i), \tag{7.16}$$

where $d(k_i,l_i)$ is a metric determining the distance between two given nodes of the network. For a geophysical transport network the geographical distance (on the sphere) between the centers of the nodes is the most natural choice. For a given pair of nodes (I,J) the average distance between the subset $\mathcal{K}_{IJ}^{M}(r,\epsilon)$ and the MPP connecting them in M time steps is defined as

$$\mathcal{D}_{IJ}^{M} = \frac{1}{N_{IJ}^{M}}\sum_{\mu}\mathbf{d}(\mu,\eta_{IJ}^{M}), \tag{7.17}$$

where N^M_{IJ} is the number of paths μ in the subset $\mathcal{K}^M_{IJ}(r, \epsilon)$, and the sum is extended over all paths in the subset (remember that η^M_{IJ} denotes the MPP). This quantity provides an estimation of how much paths in the subset deviate spatially from the corresponding MPP. A large deviation means that the probability to reach J from I is spatially spread in a large region and indicates furthermore the importance of considering the HPP subset instead of only the MPP. Small values of $\mathcal{D}^M_{IJ}$ imply HPP sets with the shape of coherent narrow tubes around the MPP, so that the MPP already characterizes the spatial pathways, even if its probability is not large.

CHAPTER 8

Advection corridors in the Mediterranean surface flow

8.1

MPPs in the ocean

We apply the previous concepts to the network describing surface water transport in the Mediterranean Sea described in Part I (Ser-Giacomi et al., 2015b).

We perform the calculations using a time step of $\tau = 10$ days, and considering M from 6 to 9 steps (i.e. in between 2 and 3 months, a time interval during which the horizontal-flow assumption remains a good approximation). We build 10 temporal networks, each one having t_0 as January 1st of each of the years available in the simulation database (2002 – 2011).

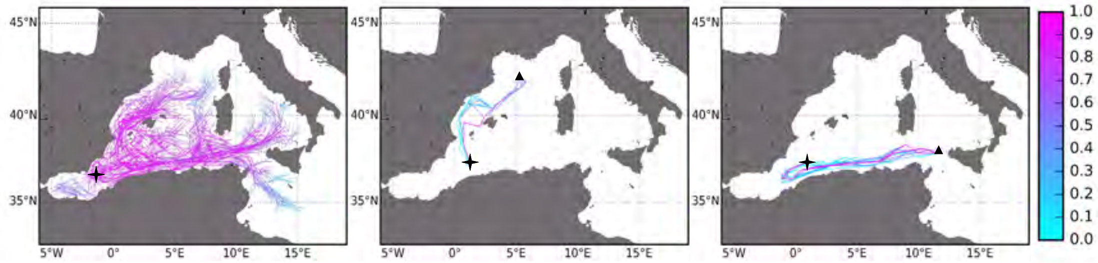

Figure 8.1: Paths of $M = 9$ steps (three months) in the Mediterranean flow network with starting date January 1st 2011, represented as straight segments joining the path nodes. Left: MPPs originating from a single node (black star) and ending in all accessible nodes. Color gives the P_{IJ}^{M} value of the paths in a normalized log-scale between the minimum value (10^{-15}, light turquoise) and the maximum (10^{-5}, dark pink). Center and right: all the paths in the $\mathcal{K}_{IJ}^{M}(\epsilon)$ set with $\epsilon = 0.1$, initial point marked by a cross and final point marked by a triangle. The center panel shows the 18 HPPs, out of a total of 54276 paths between the two sites. The MPP, with $P_{IJ}^{M} = 3 \times 10^{-9}$, is displayed in dark pink, whereas the other paths are colored with a normalized logarithmic scale according to their $(p_{IJ}^{M})_{\mu}$ values in $[\epsilon P_{IJ}^{M}, P_{IJ}^{M}]$. Right panel shows the 39 HPPs, out of a total of 61×10^{6}, in a similar logarithmic scale normalized in $[\epsilon P_{IJ}^{M}, P_{IJ}^{M}]$ with $P_{IJ}^{M} = 1.4 \times 10^{-6}$.

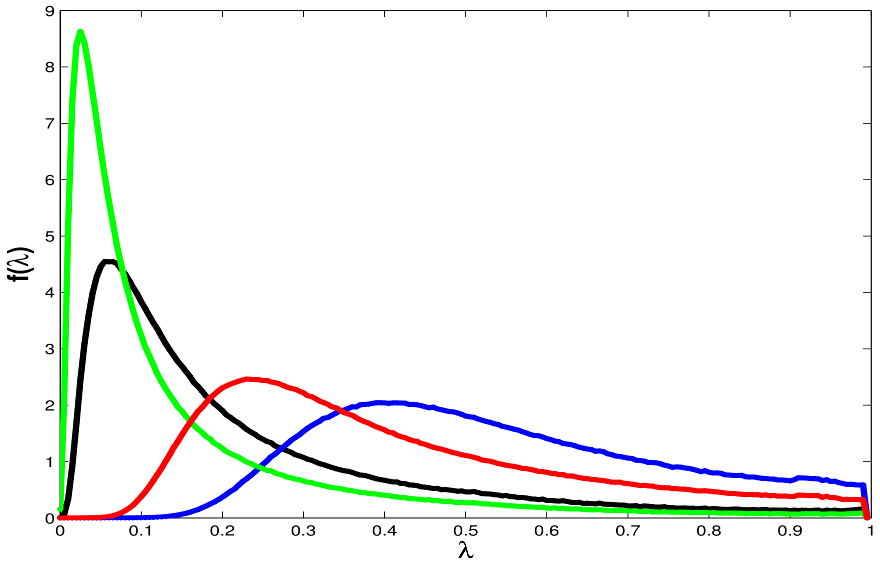

Figure 8.2: Normalized histogram $f(\lambda)$ of coefficients λ_{IJ}^M for $M = 5$ (black curve) and $M = 6$ (green curve), and $\lambda_{IJ}^M(\epsilon)$, with $\epsilon = 0.1$ for $M = 5$ (blue curve) and $M = 6$ (red curve). The statistics is compiled from all connected pairs of nodes (I, J) and the ten temporal flow networks corresponding to the first months of the ten years of velocity data. The mean values are: $\langle\lambda_{IJ}^5\rangle = 0.24$; $\langle\lambda_{IJ}^6\rangle = 0.16$; $\langle\lambda_{IJ}^5(0.1)\rangle = 0.52$; $\langle\lambda_{IJ}^6(0.1)\rangle = 0.42$.

In Fig. 8.1 (left panel) we show on map the set of all the MPPs of $M = 9$ time steps starting from a given node in one of our temporal networks (the one corresponding to 2011), and we notice how many different connections are possible from a single starting node. The P_{IJ}^M values span several orders of magnitude and this behavior is typical for the distribution of probability across MPPs. We stress here that MPPs do not coincide in general with fastest paths: the fastest connection among two nodes is not always the most probable one stressing the importance of a weighted description of the network.

To assess how representative of the whole dynamics are MPPs such as the ones shown in Fig. 8.1 we show in Fig. 8.2 the distribution of λ_{IJ}^M and $\lambda_{IJ}^M(\epsilon)$ for two values of M (see Eq. (7.14) and (7.15)). The distributions are collected from the λ-values of the whole set of accessible pairs (I, J) in our ten temporal networks. For small M most of the MPP have significant λ-values, but as M increases the peak in the distribution of λ_{IJ}^M shifts towards zero (we have checked that exponentially) as a consequence of the dramatic increase with M of the number of available paths between two nodes. Then, it becomes important to consider larger sets of paths such as $\mathcal{K}_{IJ}^M(\epsilon)$. For the cases plotted, i.e. $M = 5, 6$ and $\epsilon = 0.1$, the mean values of $\lambda_{IJ}^M(\epsilon)$ are around 0.5. This means that, despite $\mathcal{K}_{IJ}^M(\epsilon)$ may not contain the full set of paths which individually carry a probability larger than ϵP_{IJ}^M, it is large enough so that, for most of the (I, J) pairs, it contains globally over 50% of the connection probability. This result further gains meaning if we consider that the number of paths in $\mathcal{K}_{IJ}^M(\epsilon)$ is on average well below 1% of the total number of paths of $M = 5$ and 6 steps. Hence, despite the strong particle dispersion characterizing our flow networks it is true that only a small subset of paths contribute significantly to the main transport features.

8.2

MPP-betweenness centrality and multistep approach

Equipped with the above definitions we can now characterize network properties that are dependent on optimal paths in different ways. One of these is the concept of *betweenness centrality*, which is generally defined as the proportion of shortest paths passing through a node. We introduce here a definition based on the number of *most probable paths* crossing a node. Specifically we define the betweenness of node K after M steps as $\mathcal{B}_K^M = \sum_{IJ} g_{IJ;K}^M / N_M$, where the sum is over all pairs of initial nodes I and final accessible nodes J, N_M is the total number of connected pairs of nodes at time step M (computable from accessibility matrices (Lentz et al., 2013)), and $g_{IJ;K}^M$ is the number of times the node K appears in the most probable path connecting I and J. The sum can be calculated considering (or not) the initial and final nodes in the sum, and neglecting (or not) counts when paths are stopping in the same region for more than one consecutive time step. In the example studied here, all these alternative definitions lead to the same final result. Fixing the time interval M corresponds to considering paths with the same temporal duration. In this way we ignore connections that are occurring at shorter or longer times (Kim and Anderson, 2012) and that can be significantly more probable. It is possible to overcome this limitation by performing a multistep analysis: we can look at all MPPs with M in a given interval $[M_{min}, M_{max}]$ and choose the MPP, $\eta_{IJ}^{[M_{min}, M_{max}]}$, with the highest probability. The multistep analysis leads to an alternative definition of betweenness, i.e a multistep MPP-betweenness $\mathcal{B}_K^{[M_{min}, M_{max}]}$ which is calculated considering the multistep MPPs instead of the fixed-M one.

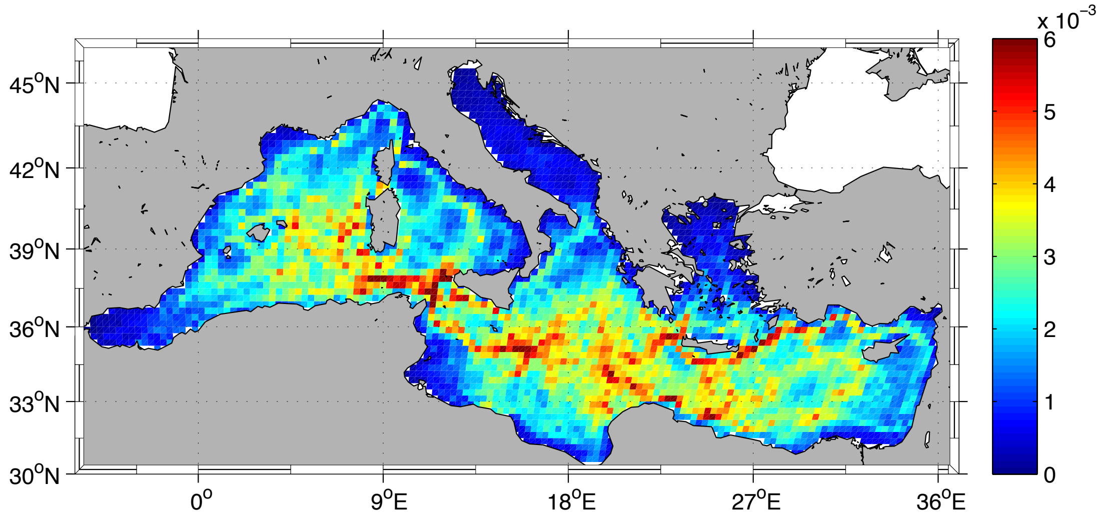

Figure 8.3: Multistep MPP-betweenness $\mathcal{B}_K^{[6,9]}$ at each geographical node K, computed for each of our ten (2002-2011) temporal networks and then averaged over them.

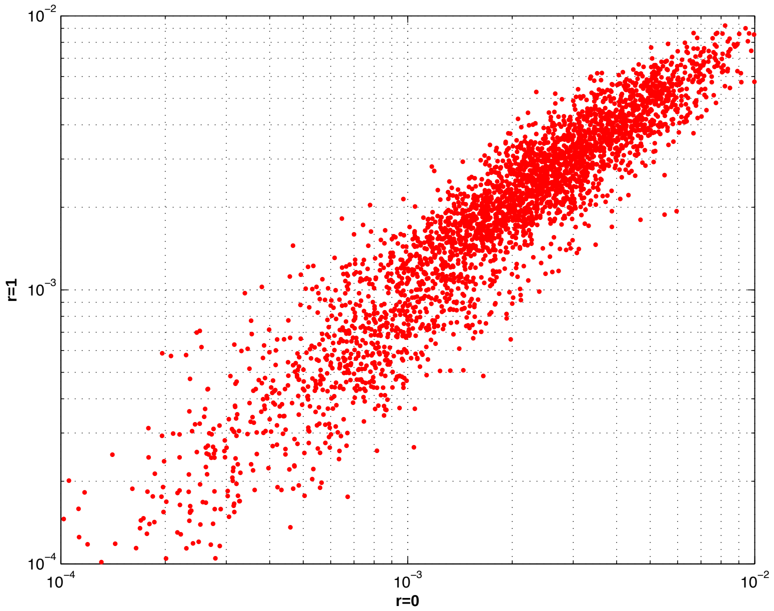

Figure 8.4: Scatter plot of MPP-betweenness (x-axis) vs. HPP-betweenness (y-axis) at 9-steps. The Pearson correlation coefficient is 0.92.
Starting date: 2011 Jan, 1.

In Fig. 8.3 we show the multistep MPP-betweenness $\mathcal{B}_K^{[6,9]}$, averaged over our ten networks. We have noticed that the distribution of the betweenness values decreases exponentially so that there are not strong hubs in the network. Spatial patterns determined by the transport dynamics of the flow are clearly evident in the figure where high betweenness areas are organized in one dimensional-like structures corresponding to the main corridors of transport, i.e. narrow pathways that connect different regions of the ocean. Main paths of the Mediterranean sea are observed like Cyprus and Rhodes Gyres, surrounding the Ionian basin, the Algerian current and those along the Sicily strait, etc. Note that because of the ten-years average, individual short lived mesoscale features (eddies and fronts) are averaged out. We also found that MPP-betweenness is highly sensible to the starting date t_0. This highlights the seasonal and inter-annual variability of the flow and justifies further our time-dependent approach.

We checked the robustness of our approach and evaluated the reliability of MPPs testing the stability of MPP-betweenness under different conditions. Dealing with temporal networks it is important to understand how much results are affected by the choice of the the single time step duration (Ribeiro et al., 2013). We checked this issue considering different M and τ values but the same total duration $M\tau$ and we obtained very similar results for the MPP-betweenness showing that it is robust under variations of the temporal resolution. We also note that the MPP-betweenness does not significantly change when considering only the 50% of MPPs with larger values of λ (i.e. when we use a threshold to retain only the most significant MPPs). Finally, to support our interpretation of most probable paths as main carriers of connectivity, we considered also MPP-betweenness using subsets of paths, so that $g_{IJ;K}^M$ is now the number of times node K appears in the set $\mathcal{K}_{IJ}^M(\epsilon)$ of HPPs between I and J, with $\epsilon = 0.1$, without noticing any qualitative or quantitative difference in the results. As shown in Fig. 8.4, values of betweenness in the two cases are very much correlated, and the Pearson correlation coefficient is larger than 0.9. Hence, despite the MPPs represent a small portion of the paths in the $\mathcal{K}_{IJ}^{M}(\epsilon)$ subsets (between $3-10\%$ for $\epsilon = 0.1$, depending on the value of M), which is itself a very small fraction of the full set of paths in the network, they seem to be representative of the main spatio-temporal structures describing the global dynamics. Indeed, center and right panels of Fig. 8.1 show that most of the relevant paths remain spatially close to the MPP. This observation is confirmed by calculations of the spatial dispersion between paths in $\mathcal{K}_{IJ}^{M}(\epsilon)$, whose average turns out to be of the order of the size of the boxes defining the nodes.

8.3

Comparing fastest and most probable paths

In the study of temporal networks, the concept of *fastest* path has been put forward as a natural extension of the *shortest path* of static networks. In our work we define and analyze a different type of relevant path which is the Most Probable Path (MPP). It is important to address the differences between most probable and fastest paths, here we illustrate them with data from the temporal network of surface flow in the Mediterranean Sea.

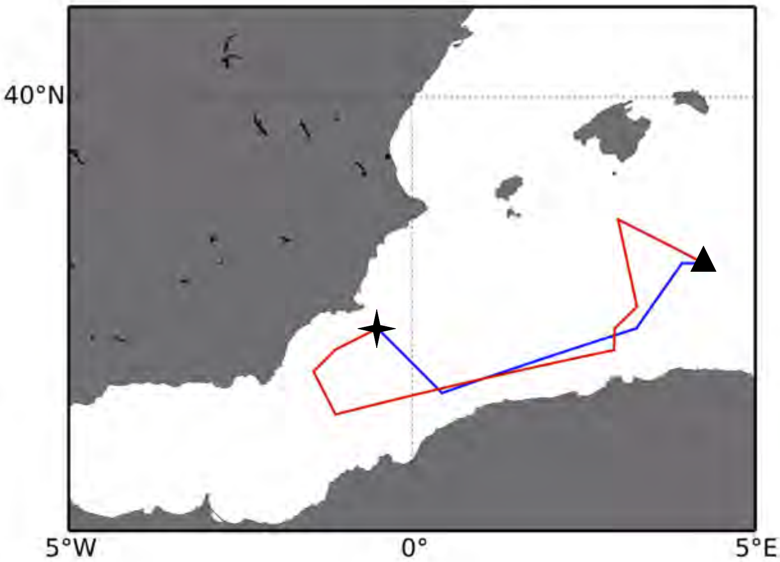

Figure 8.5: We show the *fastest*-MPP (blue) and the *absolute*-MPP (red), between an origin node I (black star) and a destination node J (black triangle). The considered full set of paths ranges The *fastest*-MPP reaches the destination node in 4 steps of $\tau = 10$ days while the *absolute*-MPP needs 8 steps i.e. 40 days more. The probability associated to the *fastest*-MPP is 5.9×10^{-7} and the probability of the *absolute*-MPP is 6.7×10^{-6}.

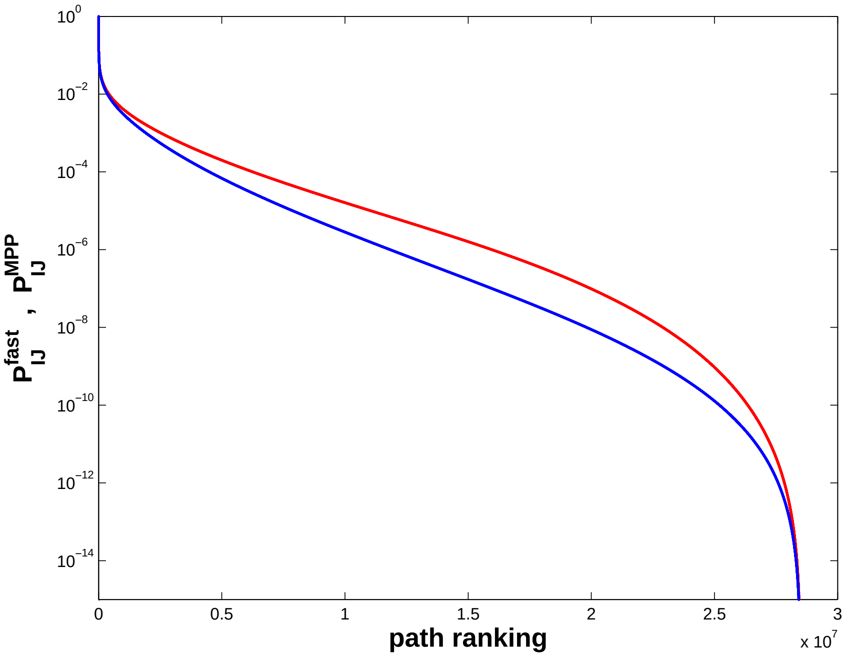

Figure 8.6: Ranking plot in which P_{IJ}^{fast} correspond to the probability of *fastest*-MPPs and P_{IJ}^{MPP} correspond to the probability of *absolute*-MPPs. The range of probability values can be read from the vertical axis (logarithmic scale). The total number of optimal paths can also be read-off from the horizontal axis.

The MPP refers to the path transporting the maximum fraction of water (or of probability) between two nodes, and the fastest path to the pathway linking the two nodes in the shortest time. This second concept can not be implemented when the number of time-steps is fixed. However we can reclaim the concept of fastest path within a multistep approach, i.e. by looking at a time window specified by a range of values for the number of time steps M. We can then define the set $\mathcal{M}_{IJ}^{[M_{min},M_{max}]}$ of $(M_{max} - M_{min} + 1)$-MPPs for the pair I, J for $M \in [M_{min}, M_{max}]$, and the *fastest*-MPP as the MPP in $\mathcal{M}_{IJ}^{[M_{min},M_{max}]}$ corresponding to the smaller M. On the other side we can also define an *absolute*-MPP, i.e. the MPP in $\mathcal{M}_{IJ}^{[M_{min},M_{max}]}$ having the highest probability. By comparing the set of *absolute*-MPPs with the set of *fastest*-MPPs we can address the question: is the fastest path necessarily the most probable?

In Fig. 8.5 we show that the *fastest*-MPP among two nodes of the network is different to the *absolute*-MPP and that its probability, in several cases, can be orders of magnitude smaller. We considered for this example paths ranging from $M = 3$ to $M = 9$ steps of 10 days (i.e. a time scale of $1 - 3$ months) with starting date January 1st 2011. The results show the importance to distinguish between the connections realized in the shortest time and the connections that carries most of the transported mass (the most probable).

To display in a more systematic way the differences between *fastest* and *absolute* MPPs across the network we study the rank plot of the whole set of paths during ten years $(2002 - 2011)$ ranging from 3 to 9 steps of 10 days starting at January 1st of each year (see Fig. 8.6). The rank plot displays of the probabilities of each path in one of the sets sorted in decreasing order. We see a gap in probabilities between the two sets of about one order of magnitude in most of the range displayed. The *fastest*-MPPs have probabilities significantly smaller than *absolute*-MPPs.

Finally, we also evaluate how these differences are reflected in the betweenness measures. We define the relative difference among the betweenness computed using the set of *fastest*-MPPs and *absolute*-MPPs for the node K as:

$$\Delta_K = 2\frac{\mathcal{B}_K^{abs} - \mathcal{B}_K^{fast}}{\mathcal{B}_K^{abs} + \mathcal{B}_K^{fast}}\,, \tag{8.1}$$

where $\mathcal{B}_K^{abs}$ is the betweenness computed using *absolute*-MPPs and $\mathcal{B}_K^{fast}$ the betweenness computed using *fastest*-MPPs. We consider once more paths ranging from 3 to 9 steps of 10 days with starting date January 1st 2011 and we compute the spatial-average value for the module of the relative difference finding $\langle|\Delta_K|\rangle_K = 0.32$. This means that, on average, the difference between the two measures is about 30%.

We stress that all the comparisons above are among paths that are already MPPs linking a pair of nodes. Considering still fastest paths (for example the one by which the very first particle from one node reaches the other) will lead to much stronger differences. In summary, the results show the importance to distinguish between the connection realized in the shortest time and the connection that carries most of the transported mass. This gains even more relevance when considering possible applications such as, rescue operations, pollutant-spreading or biological connectivity.

We considered paths in a temporal flow network describing surface flow in the Mediterranean Sea. We quantifed the relative importance of the most probable path between two nodes with respect to the whole set of paths, and to a subset of highly probable paths which incorporate most of the connection probability. Despite MPPs represent only a small fraction of the whole set of paths, we found that they suffice to highlight the main transport pathways across our network. We provide an alternative definition of betweenness centrality. High betweenness areas are organized in one dimensional-like narrow pathways that connect different regions of the ocean. Finally we showed the difference among *fastest* paths and MPPs stressing the different physical interpretation of such paths in temporal networks.

8.4

Conclusions

To conclude, we considered paths in a temporal flow network describing surface flow in the Mediterranean Sea. We quantified the relative importance of the most probable path between two nodes with respect to the whole set of paths, and to a subset of highly probable paths which incorporate most of the connection probability. Despite MPPs represent only a small fraction of the whole set of paths, we found that they suffice to highlight the main transport pathways across our network. We provide an alternative definition of betweenness centrality. High betweenness areas are organized in one dimensional-like narrow pathways that connect different regions of the ocean. Finally we showed the difference among *fastest* paths and MPPs stressing the different physical interpretation of such paths in temporal networks.

CHAPTER 9

Transport pathways in an atmospheric blocking event

9.1

Introduction

Eastern Europe and Western Russia experienced a strong, unpredicted, heat wave during the summer of 2010. Extreme temperatures resulted in over 50000 deaths and inflicting large economic losses to Russia. The heat wave was due to a strong atmospheric blocking that persisted over the Euro-Russian region from late June to early August (Matsueda, 2011). During July the daily temperatures were near or above record levels and the event covered Western Russia, Belarus, Ukraine, and the Baltic nations. Physically, the origins of this heat wave were in an atmospheric blocking episode that produced anomalously stable anticyclonic conditions, redirecting the trajectories of migrating cyclones. Atmospheric blockings can remain in place for several days (sometimes even weeks) and are of large scale (typically larger than 2000 km). In particular, the Russian blocking of summer 2010 was morphologically of the type known as *Omega blocking* that consists in a combination of low-high-low pressure fields with geopotential lines resembling the Greek letter Ω (see Fig.9.1). Omega blockings bring warmer and drier conditions to the areas that they impact and colder, wetter conditions in the upstream and downstream (Black et al., 2004). Despite these type of events have been well-investigated over the years, a complete understanding and prediction is still missing.

Here we present a characterization of this flow pattern based on the study of fluid transport as a Lagrangian flow network. The most probable paths linking nodes of this atmospheric network reveal the dominant pathways traced by atmospheric fluid particles. We study the concrete period extended from the July the 20th to July 30th (Ser-Giacomi et al., 2015c).

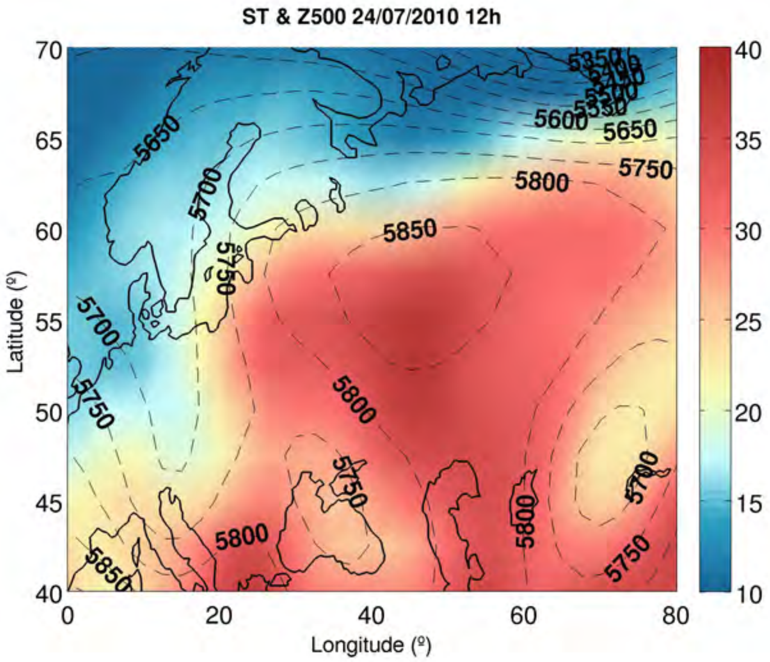

Figure 9.1: Geopotential height at 500hPa (contours, in m) and temperature (color code, in degrees C) on July 24th, 12:00 UTC.

9.2

Atmospheric setup

Atmospheric data were provided by the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) through the Global Forecast System (GFS) (Saha and coauthors, 2010). This reanalysis was initially completed over the 31 year period from 1979 to 2009 and extended to March 2011. Data can be obtained with a temporal resolution of 1 hour and a spatial horizontal resolution of $0.5° \times 0.5°$. The spatial coverage contains a range of longitudes of $0°E$ to $359.5°E$ and latitudes of $90°S$ to $90°N$. The variables needed as input to the Lagrangian dispersion model described in the next section include dew point temperature, geopotential height, land cover, planetary boundary layer height, pressure and pressure reduced to mean sea level, relative humidity, temperature, zonal and meridional component of the wind, vertical velocity and water equivalent to accumulated snow depth. All these fields are provided by CFSR data on 26 pressure levels.

As mentioned, the idea is to obtain the effective velocity field felt by any fluid particle. Then the Lagrangian dispersion model (see next subsection) will integrate it to provide as output the three-dimensional positions of the particle at every time step. The numerical model used to integrate particle velocities and obtain trajectories is the Lagrangian particle dispersion model FLEXPART version 8.2 (Stohl et al., 2005, 2011). FLEXPART simulates the long-range and mesoscale transport, diffusion, dry and wet deposition, and radioactive decay of tracers released from point, line, area or volume sources. It most commonly uses meteorological input fields from the numerical weather prediction model of the European Centre for Medium-Range Weather Forecasts (ECMWF) as well

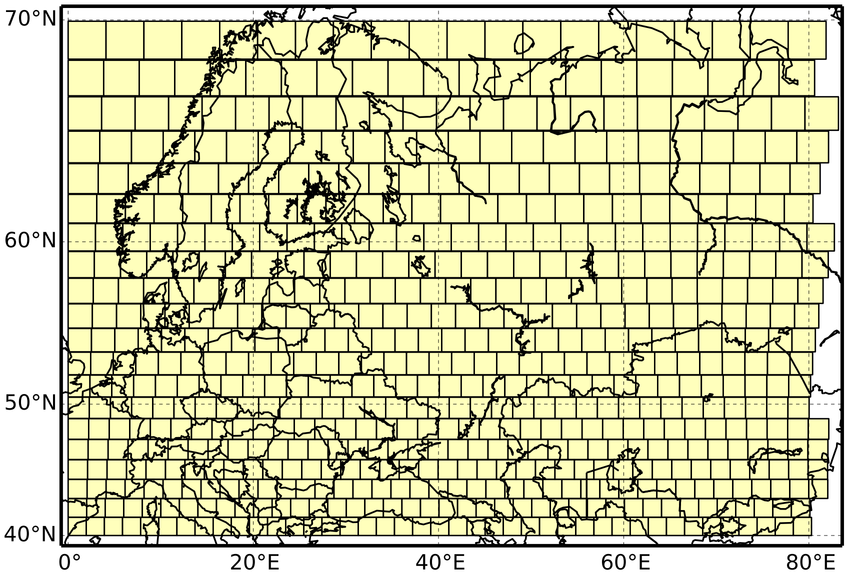

Figure 9.2: The geographical domain considered and the discretization grid defining the nodes of our flow network.

as the Global Forecast System (GFS) from NCEP (the one used in our study). Trajectories are produced by integrating the equation (the input velocity data are interpolated on the present particle position):

$$\frac{d\mathbf{X}}{dt} = \mathbf{v}(\mathbf{X}(t)), \tag{9.1}$$

with t being time, $\mathbf{X}$ the vector position of the air particle, and $\mathbf{v} = \bar{\mathbf{v}} + \mathbf{v}^{\mathbf{t}} + \mathbf{v}^{\mathbf{m}}$ is the wind vector. FLEXPART takes the grid scale wind $\bar{\mathbf{v}}$ from the CFSR, but complements it with stochastic components $\mathbf{v}^{\mathbf{t}}$ and $\mathbf{v}^{\mathbf{m}}$ to better simulate the unresolved turbulent processes occurring at small scales. The turbulent wind fluctuations $\mathbf{v}^{\mathbf{t}}$ are parametrized by assuming a Markov process via a Langevin equation, and the mesoscale wind fluctuations $\mathbf{v}^{\mathbf{m}}$ are implemented also via an independent Langevin equation by assuming that the variance of the wind at the grid scale provides information on the subgrid variance. Variables entering the parametrizations are obtained from the meteorological CFSR fields. For additional details we refer to (Stohl et al., 2005, 2011).

We focus our analysis on the domain in between 0°E - 80°E and 40°N - 70°N. In order to define the nodes of the network we discretize this region in 626 equal-area boxes using a sinusoidal projection. The latitudinal extension of each node-box is 1.5°, the longitudinal one varies depending on the latitude (see Fig. 9.2). The area of each box is 27722 km^2, so that the typical horizontal size is of the order of 166.5 km. We take $\tau = 12$ hours as time discretization. We uniformly fill each node with 800 ideal fluid particles releasing them at 5000 m of height, a representative level in the middle troposphere. FLEXPART trajectories are fully three-dimensional, but by initializing at each time-step particles in a single layer we are effectively neglecting the vertical dispersion (which is of the order of 800 m in the $\tau = 12\ h$ time step) and focussing on the pathways of large scale horizontal transport. Fully three-dimensional flow networks will be the subject of future work.

9.3

MPPs in the atmosphere

Equipped with the tools developed above we can now compute pathways of transport during the atmospheric event described above. Figure 9.3a shows all the optimal paths leaving a node in the Scandinavian Peninsula at July 25 and arriving to all nodes which are reached in $M = 9$ steps (i.e. 4.5 days).

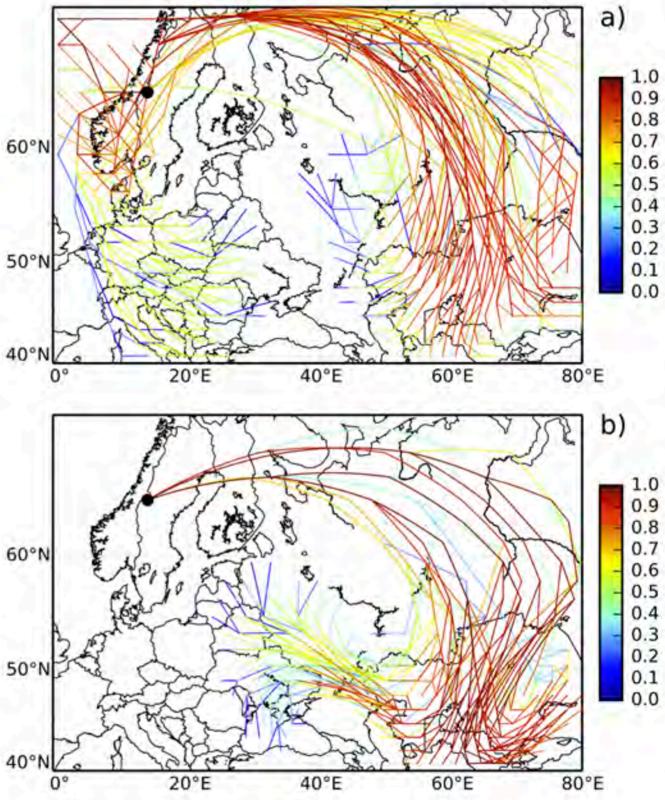

Figure 9.3: Paths of M = 9 steps of $\tau = 12$ hours in our flow network with starting date July 25th 2010 (panel a)) and July 20th 2010 (panel b)), represented as geodesic arcs joining the path nodes. MPPs originating from a single node (black circle) and ending in all accessible nodes. Color gives the P^M_{IJ} value of the paths in a normalized log-scale between the minimum value (deep blue) and the maximum (dark red). Panel a): probabilities ranging from 10^{-3} to 10^{-14}. Panel b): probabilities ranging from 10^{-3} to 10^{-15}.

The graphical representation joins with maximal arcs the center of the grid boxes identified as pertaining to the MPP. The actual particle trajectories between two consecutive boxes are not necessarily such arcs. The paths are colored according to their probability value P_{IJ}^{M}.

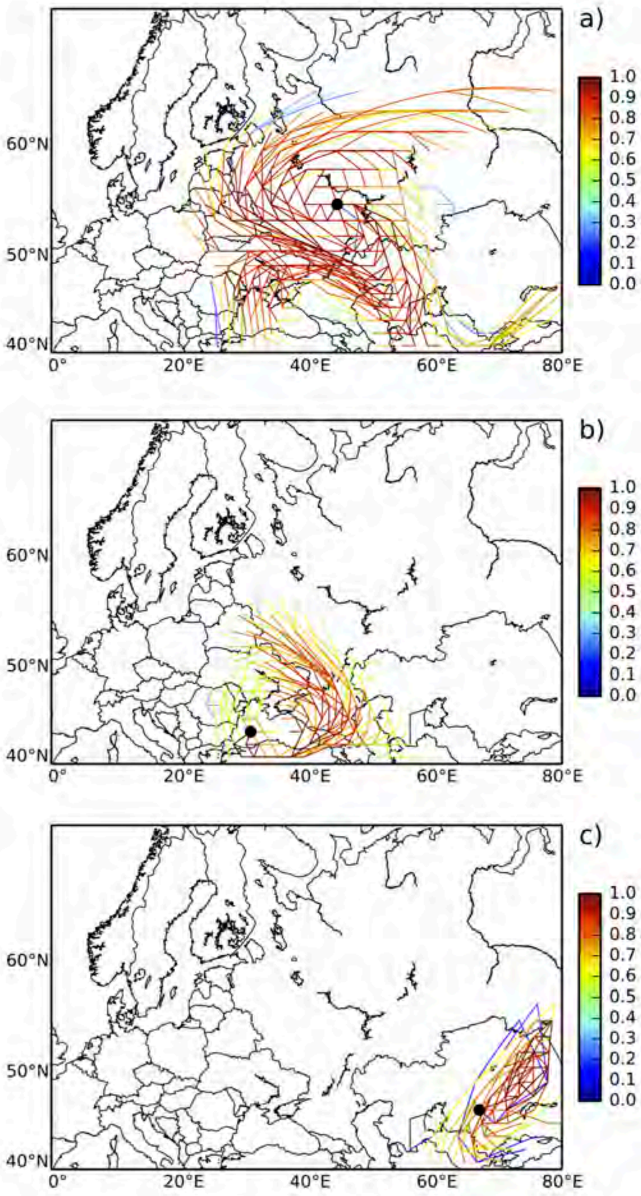

Figure 9.4: Optimal paths of 9 steps of $\tau = 12$ hours with starting date July 20th 2010, entrained in the high- and in the two low-pressure areas of the blocking. Same coloring scheme as in Fig. 9.3. Panel a): probabilities ranging from 10^{-3} to 10^{-16}. Panel b): probabilities ranging from 10^{-2} to 10^{-16}. Panel c): probabilities ranging from 10^{-3} to 10^{-13}.

The MPPs with highest probability (reddish colors) follow a dominant anticyclonic (i.e. clockwise) route bordering the high pressure region (see Fig. 9.1, but note that this is at a particular time, whereas the trajectory plots span a range of dates of more than four days) without penetrating it. There is also a branch of MPPs with much smaller probabilities (yellow and bluish colors) that are entrained southward by a cyclonic circulation.

Despite the persistent character of the Eulerian block configuration, sets of Lagrangian trajectories become highly variable in time. See for example the set of MPPs starting from the same initial location but five days earlier (Fig. 9.3b). The southward cyclonic branch is now absent, all MPPs following initially the anticyclonic gyre. Remarkably, the set of trajectories bifurcates into two branches when approaching what seems to be a strong hyperbolic structure close to 40°N 60°E. A hint of the presence of second hyperbolic structure is visible at the end of the westward branch, close to 50°N 30°E. Figure 9.4 displays additional MPPs starting also at July 20th, but initialized inside the main anticyclonic region of the blocking, and in two low-pressure regions flanking it. Fig. 9.4a clearly shows the main anticyclonic circulation, highlighting also the escape routes from the high-pressure zone, associated with the hyperbolic regions described above. The other two panels show the cyclonic circulations at each side of the high, in a characteristic Omega-blocking configuration. It is remarkable the compactness of the trajectories inside the eastern low-pressure area, which form a very localized and coherent set with practically no escape in the 4.5 days time-interval displayed.

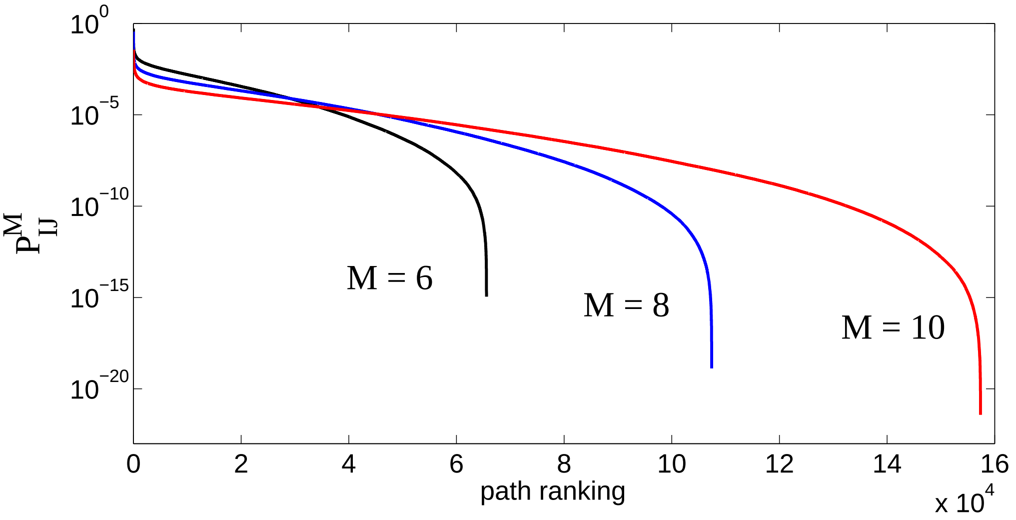

Figure 9.5: Ranking plot in which the P^M_{IJ} values of all MPPs obtained for $M = 6{,}8$, and 10 starting on July 25th in the whole area are plotted in decreasing order. The range of probability values of the MPPs can be read from the vertical axis (from a few percent to 10^{-15} for $M = 6$ or to less than 10^{20} for $M = 10$). The total number of optimal paths can also be read-off from the horizontal axis.

We stress that the plots in Figs. 9.3 and 9.4 are different from *spaghetti plots* for which many available trajectories are plotted from different or related initial conditions. For our set of particles this will give 800 trajectories emanating from each box. Here we are plotting just one path, the MPP, for each initial and final box pair, which strongly limits the number of paths from each box but, as we will see more thoroughly, it is still representative of the trajectories of many released particles.

9.4

Relevance of MPPs

The range of colors in Figs. 9.3 and 9.4 indicates that, given an initial box, not all MPPs leading to different locations are equally probable. This is quantified by the probability P_{IJ}^M which gives a weight to each MPP. Indeed P_{IJ}^M takes a very large range of values.

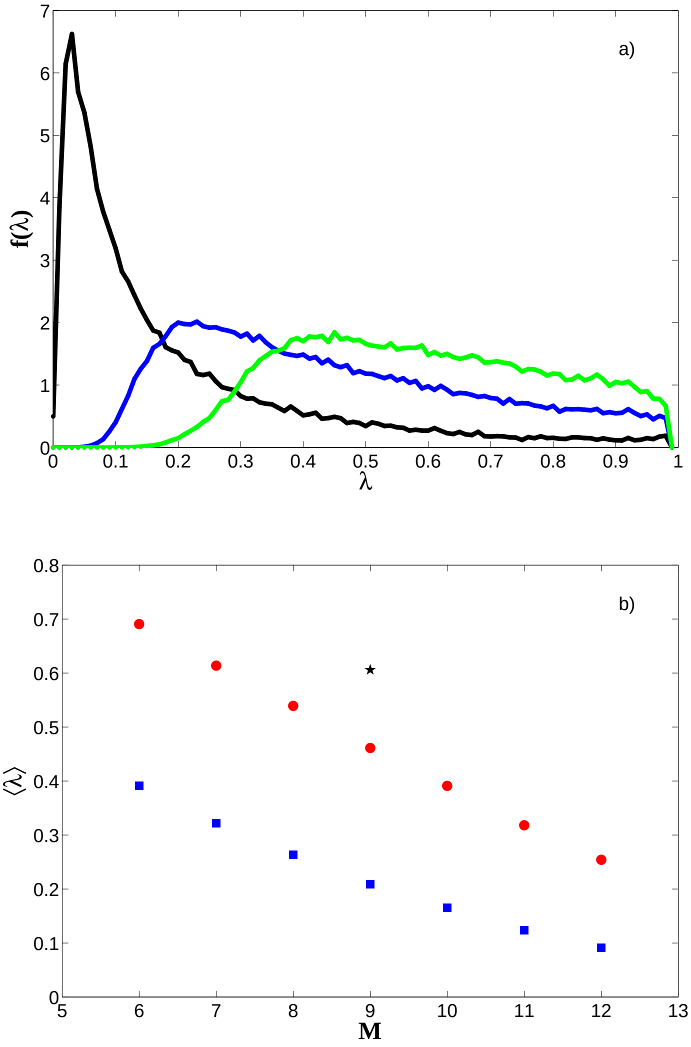

Figure 9.6: a) Normalized probability density $f(\lambda)$ of the merit figure $\lambda_{IJ}^M(r,\epsilon)$ of paths started on July 20th 2010 for M = 9 and $\epsilon = 0.1$, with $r = 0$ (only the MPPs, black curve), $r = 1$ (blue) and $r = 2$ (green). b) Mean value of the $\lambda_{IJ}^M(r,\epsilon)$ distributions (paths' starting date July 25th) as a function of the number of time steps M for $r = 0$ (only MPPs, blue squares), $r = 1$ (red circles) and $r = 2$ (single black star).

Figure 9.5 shows a ranking plot in which the values of all MPPs of a given M and starting at a particular date are plotted in decreasing order. We see a huge spread on the values of P_{IJ}^M. Very low probability values arise because of the exponential explosion

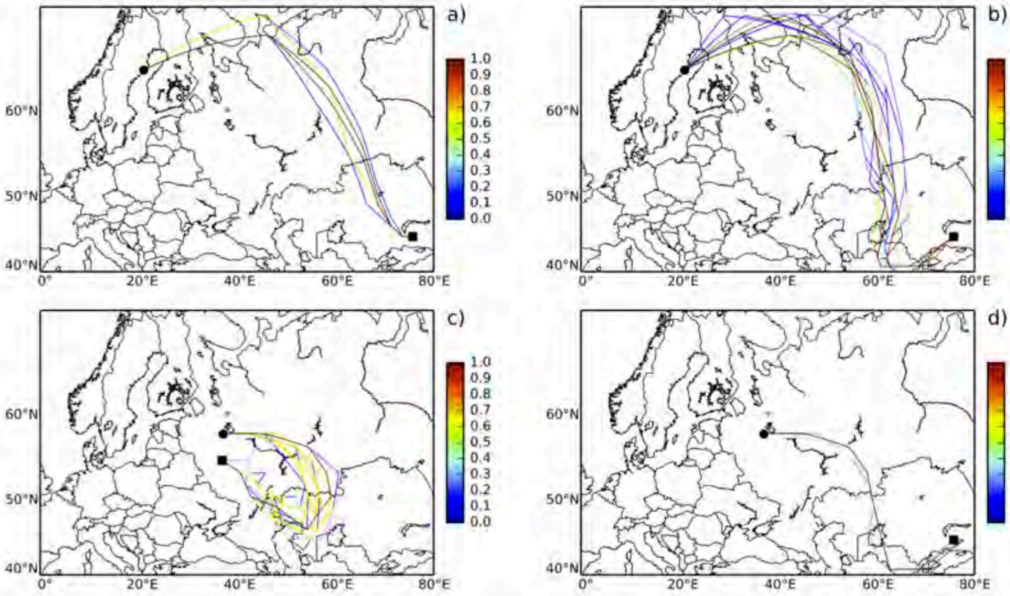

Figure 9.7: All paths in $\mathcal{K}^M_{IJ}(r = 1, \epsilon = 0.1)$ for different I, J, initial point I marked by a circle and final point J marked by a square. Panel a): M=8 steps, with starting date July 25th 2011. Panel b): M=12 steps, with starting date July 25th 2011. Panel c): M=11 steps, with starting date July 20th 2011. Panel d): M=11 steps, with starting date July 20th 2011. Same coloring scheme as in Figs. 9.3 and 9.4. For panels a) and b) probabilities ranging from 10^{-5} to 10^{-6} while for panels c) and d) from 10^{-7} to 10^{-8}. The maximum probability in each panel (dark red) corresponds to the MPP.

of the number of paths between two nodes with increasing M. Given these low values of P^M_{IJ} except for the smallest values of M, one should ask how representative are the MPPs for the full set of paths. Figure 9.6a shows distributions of the parameter $\lambda^M_{IJ}(r, \epsilon)$ giving the relative importance of the different types of paths (see Chapter 7). We see that λ-values are small when considering only the MPPs ($r = 0$), but the distributions shift towards higher values for HPPs sets of increasing r. Figure 9.6b gives mean values of the λ distributions. They decrease with M, reflecting the lack of representativeness of the smallest sets of paths for large M. However, already for $r = 1$ the set of HPPs has a mean value higher than 0.5 for a relevant range of time steps.

Thus, for the values of M and ϵ discussed here, the set of HPPs with $r = 1$ seems to be rich enough to represent the transport pathways. But how different is the geometry of the different paths in this HPP set? And how different is it from the MPPs? We plot in Fig. 9.7 examples of all HPPs with $r = 1$ and $\epsilon = 0$ for particular (I, J) values and dates. Changing M changes the shape of the path, but in all the cases the sets remain coherent and narrow tubes of trajectories defining roughly the same pathway as the MPP.

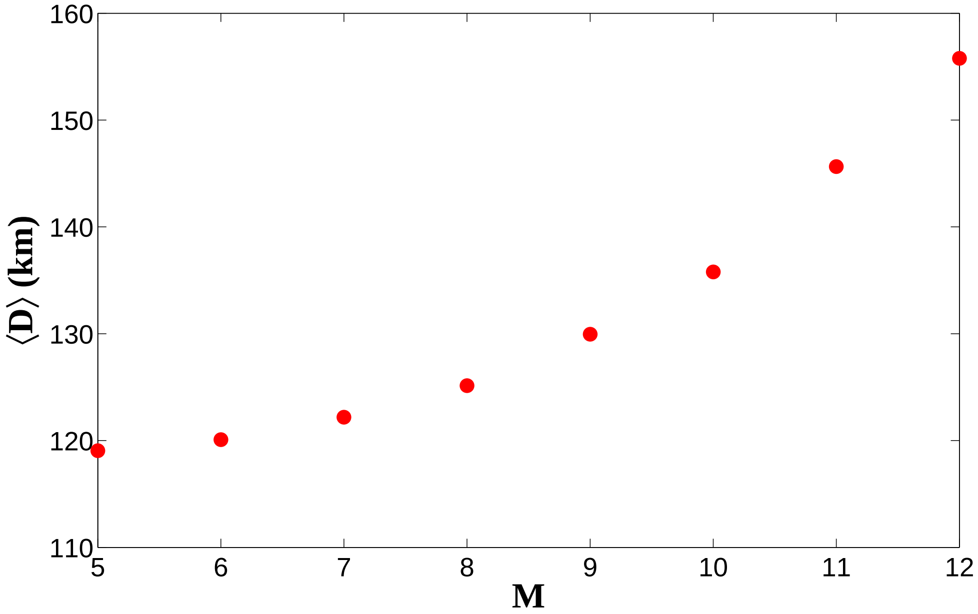

Figure 9.8: Plot of the mean distance $\mathcal{D}_{IJ}^{M}$ (Eq. (7.17)) as a function of M for $r = 1$ and $\epsilon = 0.1$. The quantity is further averaged over all the HPPs starting on July 25th. Units are kilometers.

A quantification of the *width* of the tubes can be done with the distance measure $\mathcal{D}_{IJ}^{M}$ in Eq. (7.17). An average of it over pairs of locations is shown in Fig. 9.8. Although the tube width increases with M, it remains always below the typical linear box size of approximately 166.5 km indicating that the tubes remain narrow. Thus we conclude that, despite the decreasing probability of the MPPs for increasing M, they remain good indicators of the dominant pathways in the transport network.

9.5

Conclusions

To conclude, we computed optimal transport paths for the atmospheric circulation during a blocking event occurring in Summer 2010 (in particular we focus our study for the period 20th July - 30th July) over Eastern Europe and Russia. This atmospheric flow has very different temporal and spatial scales, and is much more diffusive, than the oceanic flow analyzed in Chapter 8. As in the previous Chapter we found that the significance of single MPPs decreases with the number of steps considered, but the MPPs remain representative of the spatial geometry of the pathways, in the sense that the sets of HPPs are coherent narrow tubes providing transport paths always close to the optimal path.

Conclusions and Perspectives

We have used the concept of flow networks to obtain a discretized view of transport processes in geophysical contexts. Once the fluid motion is cast into the graph theory framework, powerful techniques from this field become available to investigate the fluid transport processes. In Part I we have improved, using network concepts, the characterization of geophysical dispersion and mixing process, as well as the identification of coherent fluid regions. One of the simplest network descriptors, the degree of a node, gives direct information on local stretching properties, classically associated to the finite-time Lyapunov exponents and their distributions. Thus the out-degree at a particular node is quantitatively related to the fluid stretching at that location in the time forward direction, and quantifies fluid dispersion. The in-degree is related to backwards stretching, and thus to the mixing of fluid from different origins (Ser-Giacomi et al., 2015a).

A family of *network-entropy* functions has been defined, aiming at describing higher-order statistical properties of fluid stretching (and then of dispersion and mixing) in terms of the network adjacency matrix. One of them, H_i^0 is simply the logarithm of the degree. Another one, H_i^1, is the discrete finite-time entropy studied by (Froyland and Padberg-Gehle, 2012). We find numerically that it provides a good estimation of the coarse-grained finite-time Lyapunov exponent (Ser-Giacomi et al., 2015a). We expect higher-order entropies to be related to the generalized Lyapunov exponents that characterize successive moments of the Lyapunov field. Further work is needed to assess the validity of these properties more rigorously.

We considered flow networks in the geophysical context, but it is anticipated that the concepts are equally valid in more general fluid dynamics context, and even apply to more abstract flows in the phase space of dynamical systems. Also, the network entropies defined here can be used to characterize the local properties of general weighted networks beyond the degree and the node strength. They represent indeed the way to define and measure the local Lyapunov exponent in any weighted network (Ser-Giacomi et al., 2015a).

As a second application in which the network representation provides useful insights we have investigated the identification of coherent regions in the ocean flow, regions that are similar to *almost invariant* sets but for which the presence of strong internal mixing is also desired. We find in the network-theory toolbox a useful community detection technique, *Infomap* (Rosvall and Bergstrom, 2008), that takes into account the directed and weighted nature of the flow network, and that finds partitions of the geophysical flow with the required characteristics without restricting the range of community sizes. We have argued that these characteristics make it an interesting alternative to spectral methods to identify the desired coherent regions, although we also recognize that a substantial clarification of the physical meaning of the minimization process involved in *Infomap* is needed (Ser-Giacomi et al., 2015a).

This partition of the sea into coherent *provinces* is used in Part II to evaluate larval connectivity and to inform the design of marine protected areas in the Mediterranean Sea. The systematic detection of communities in the network extracts a set of hydrodynamical provinces which organize the dispersion of larvae in the entire Mediterranean basin. Their boundaries coincide with both mesoscale and regional-scale oceanographic features, comprehending the multi-scale processes of ocean circulation. The repeated occurrence of these frontiers allows separating the seascape into different hydrodynamical units which provide the "backbone" of oceanic transport impacting larval dispersal and connectivity within existing MPAs. While the role of such large-scale dispersal patterns on the genetic structure of marine population remains to be determined, the hydrodynamical units evidenced may be used to optimize the sampling strategy for genetic studies (Rossi et al., 2014).

Various connectivity metrics and their spatiotemporal variability are then related to local oceanography. We show that retention processes are favored along coastlines with sluggish circulation and over extended continental shelves. We also demonstrate that convergence/divergence areas generated by alongshore winds are often associated with larval sinks/sources. Furthermore, studying the relationship among the connectivity metrics sheds light on their integrated interpretation for conservation planning. Overall, our results offer many opportunities such as (i) improving the sampling strategy of genetic studies, (ii) formulating and testing hypotheses about the role of larval dispersal in structuring spatially and genetically populations and (iii) depicting systematically retention and transport processes which should help scientists and managers to optimize the implementation of future MPAs, ultimately consolidating the present network of reserves in the Mediterranean Sea, and elsewhere in the global ocean, to achieve a sound management of marine ecosystems (Dubois et al., 2015).

In Part III we introduced tools to compute highly probable paths in weighted temporal networks and to evaluate their relative importance. Until this point we did not need any assumptions about the memory of the transport processes across the network, but from here we assumed that such precesses are Markovian. Betweenness centrality measures based on them have also been introduced. First we applied this approach to characterize connectivity in the Mediterranean Sea from a network-theory perspective. Here, MPPs correspond also to the set of paths that maximize the fraction of transported mass, giving therefore a clear physical interpretation of connection probabilities. Despite MPPs represent only a small fraction of the whole set of paths, we found that they suffice to highlight the main transport pathways across our oceanic network, since most of the HPPs remain geographically close to them (Ser-Giacomi et al., 2015b). Finally we computed MPPs and HPPs for an atmospheric blocking event involving eastern Europe and Western Russia. The computed optimal paths give a Lagrangian view of the Omega-block configuration, with a central anti-cyclonic circulation flanked by two cyclonic ones. Moreover they give additional insight on it, such as the variability of the dominant pathways, and the identification of escaping and trapping regions. The statistical significance of single MPPs decreases with the time interval considered, but, as for the case of the Mediterranean, we find always that the MPPs remain representative of the spatial geometry of the pathways, in the sense that the sets of HPPs are coherent narrow tubes providing transport paths always close to the optimal path (Ser-Giacomi et al., 2015c). This seems to be thus a general characteristic of geophysical flow networks.

We believe that the study of fluid transport from the point of view of MPPs connectivity will provide new tools and insights complementing standard Lagrangian methods. Indeed most of these are devoted to the identification of barriers to transport or coherent regions with small fluid exchange with the surroundings. Here we are instead addressing the opposite question: how to detect regions and pathways that maximize fluid interchange across the network. Even though in principle pathways are simply given by trajectories, it is almost impossible to extract clear and significant patterns from the complex sets of trajectories that arise in all, except the most simple, time-dependent flows. Our approach allows to quantify explicitly transport among two sub-regions of the domain, highlighting the optimal path connecting them. In this sense MPP-analysis differs from simply studying the evolution in time of tracer concentrations seeded in a given region (Ser-Giacomi et al., 2015b,c).

We believe that the representation of fluid motion as a transport or flow network, allowing the use of powerful techniques from graph theory, will continue to provide novel insights into the nonlinear processes occurring in our planet, most of them related to fluid transport.

The present implementation of the method deals only with regions fixed with respect to geographic boundaries. Tools from the study of bipartite networks would be needed to find moving coherent regions such as vortices and eddies.

The similarity between our flow-driven boundaries and major environmental gradients commonly used to regionalize the Mediterranean seascape finally suggests they might also define oceanic biomes or even faunistic units. Future developments may help incorporating large-scale biogeography and dispersal patterns to improve MPAs design for efficient management and conservation of marine ecosystems. Future developments of our discretized model are also envisaged to resolve additional factors influencing connectivity. Abiotic variables may affect many biological parameters. Realistic larval traits such as an active swimming could be also implemented, but only large vertical migrations are expected to change significantly the dispersal schemes. Concerning our spatial discretization, on- going work indicates that smaller nodes return finer spatial structures but do not change the patterns nor the magnitude of our connectivity proxies.

We believe that the study of fluid transport as a network will provide new tools and insights complementing standard Lagrangian methods. Beyond the fluid dynamics context, MPPs and the MPP-betweenness measure here introduced could be easily transferred to other kinds of weighted temporal networks. This could be relevant, for instance, in defining vulnerability metrics in disease spreading processes, or in detecting bottlenecks of reaction chains in metabolic networks.

Acknowledgments

I want to thank my two supervisors Emilio Hernández-García and Cristóbal Lopez for giving me the great opportunity of writing this thesis and for supporting me from the beginning to the end of this work.

I acknowledge also financial support from FEDER and MINECO (Spain) through the ESCOLA (CTM2012- 39025-C02-01) and INTENSE@COSYP (FIS2012-30634) projects, and from European Commission Marie-Curie ITN program (FP7-320 PEOPLE-2011-ITN) through the LINC project (no. 289447).

Bibliography

Aldecoa, R. and Marín, I. (2013). Exploring the limits of community detection strategies in complex networks. *Scientific Reports*, 3:2216.

Anadón, J., Mancha-Cisneros, M., Best, B., and Gerber, L. (2013). Habitat-specific larval dispersal and marine connectivity: implications for spatial conservation planning. *Ecosphere*, 4(7):art82.

Andrello, M., Mouillot, D., Beuvier, J., Albouy, C., Thuiller, W., and Manel, S. (2013). Low connectivity between Mediterranean marine protected areas: a biophysical modeling approach for the dusky grouper *epinephelus marginatus*. *PloS one*, 8(7):e68564.

Aurell, E., Boffetta, G., Crisanti, A., Paladin, G., and Vulpiani, A. (1997). Predictability in the large: an extension of the Lyapunov exponent. *J. Phys. A*, 30:1–26.

Bakun, A. and Agostini, V. (2001). Seasonal patterns of wind-induced upwelling/downwelling in the Mediterranean sea. *Scientia Marina*, 65(3):243–257.

Banks, S., Piggott, M., Williamson, J., Bové, U., Holbrook, N., and Beheregaray, L. (2007). Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology*, 88(12):3055–3064.

Barreira, L. and Pesin, Y. B. (2002). *Lyapunov exponents and smooth ergodic theory*, volume 23. American Mathematical Soc.

Bettencourt, J. P. H. (2014). *Three dimensional lagrangian structures in turbulent flows: application to oceanic processes*. PhD Thesis. `https://ifisc.uib-csic.es/publications/publication-detail.php?indice=2575`.

Black, E., Blackburn, M., Harrison, G., Hoskins, B., and Methven, J. (2004). Factors contributing to the summer 2003 European heatwave. *Weather*, 59:217–223.

Bode, M., Bode, L., and Armsworth, P. (2006). Larval dispersal reveals regional sources and sinks in Great Barrier Reef. *Marine Ecology Progress Series*, 308:17–25.

Boffetta, G., Cencini, M., Falcioni, M., and Vulpiani, A. (2002). Predictability: a way to characterize complexity. *Physics Reports*, 356(6):367 – 474.

Botsford, L., Brumbaugh, D., Grimes, C., Kellner, J., Largier, J., O'Farrell, M., Ralston, S., Soulanille, E., and Wespestad, V. (2009). Connectivity, sustainability, and yield: bridging the gap between conventional fisheries management and marine protected areas. *Reviews in Fish Biology and Fisheries*, 19(1):69–95.

Bouffard, J., Pascual, A., Ruiz, S., Faugère, Y., and Tintore, J. (2010). Coastal and mesoscale dynamics characterization using altimetry and gliders: A case study in the Balearic sea. *J. Geophys. Res.*, 115:C10029.

Branicki, M. and Wiggins, S. (2010). Finite-time Lagrangian transport analysis: stable and unstable manifolds of hyperbolic trajectories and finite-time Lyapunov exponents. *Nonlinear Processes in Geophysics*, 17(1):1–36.

Burgess, S., Nickols, K., Griesemer, C., Barnett, L., Dedrick, A., Satterthwaite, E., Yamane, L., Morgan, S., White, J., and Botsford, L. (2014). Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design. *Ecological Applications*, 24(2):257–270.

Caldarelli, G. (2007). *Scale-free networks: complex webs in nature and technology*. Oxford University Press.

Calò, A., Félix-Hackradt, F., Garcia, J., Hackradt, C., Rocklin, D., Treviño Otón, J., and García-Charton, J. (2013). A review of methods to assess connectivity and dispersal between fish populations in the Mediterranean sea. *Advances in Oceanography and Limnology*, 4(2):150–175.

Carson, H., Cook, G., López-Duarte, P., and Levin, L. (2011). Evaluating the importance of demographic connectivity in a marine metapopulation. *Ecology*, 92(10):1972–1984.

Castiglione, P., Falcioni, M., Lesne, A., and Vulpiani, A. (2010). *Chaos and Coarse Graining in Statistical Mechanics*. Cambridge University Press, Cambridge.

Cencini, M., Cecconi, F., and Vulpiani, A. (2010). *Chaos: From simple models to complex systems*. World Scientific, Singapore.

Chaoui, L., Kara, M., Quignard, J., Faure, E., and Bonhomme, F. (2009). Forte différenciation génétique de la daurade *Sparus aurata* entre les deux rives de la méditerranée occidentale. *Comptes Rendus Biologies*, 332(4):329–335.

Coll, M., Cury, P., Azzurro, E., Bariche, M., Bayadas, G., Bellido, J. M., Chaboud, C., Claudet, J., El-Sayed, A. F., D., D. G., Knittweis, L., Pipitone, C., Samuel-Rhoads, Y., Taleb, S., Tudela, S., and Valls, A. (2013). The scientific strategy needed to promote a regional ecosystem-based approach to fisheries in the Mediterranean and Black Seas. *Rev. Fish Biol. Fisheries*, 23:415–434.

Coll, M., Piroddi, C., Albouy, C., Lasram, F., Cheung, W., Christensen, V., Karpouzi, V., Guilhaumon, F., Mouillot, D., Paleczny, M., Palomares, M., Steenbeek, J., Trujillo, P., Watson, R., and Pauly, D. (2012). The Mediterranean Sea under siege: spatial overlap between marine biodiversity, cumulative threats and marine reserves. *Global Ecol. Biogeogr.*, 21:465–480.

Coll, M., Steenbeek, J., Lasram, F., Mouillot, D., and Cury, P. (2015). Low-hanging fruit for conservation of marine vertebrate species at risk in the Mediterranean sea. *Global Ecology and Biogeography*, 24:226–239.

Colleter, M., Gascuel, D., Ecoutin, J.-M., and de Morais, L. T. (2012). Modelling trophic flows in ecosystems to assess the efficiency of marine protected areas: a case study on the coast of Senegal. *Ecol. Model.*, 232:1–13.

Corell, H., Moksnes, P.-O., Engqvist, A., Doos, K., and Jonsson, P. (2012). Depth distribution of larvae critically affects their dispersal and the efficiency of marine protected areas. *Mar. Ecol. Prog. Ser.*, 467:29–46.

Cowen, R. K., Paris, C. B., and Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science*, 311:522–527.

Cowen, R. K. and Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annu. Rev. Mar. Sci.*, 1:433–466.

Crowder, L., Lyman, S., Figueira, W., and Priddy, J. (2000). Source-sink population dynamics and the problem of siting marine reserves. *Bulletin of Marine Science*, 66(3):799–820.

Danon, L., Díaz-Guilera, A., Duch, J., and Arenas, A. (2005). Comparing community structure identification. *J. Stat. Mech.*, 09:P09008.

Dellnitz, M., Froyland, G., Horenkamp, C., Padberg-Gehle, K., and Sen Gupta, A. (2009). Seasonal variability of the subpolar gyres in the Southern Ocean: a numerical investigation based on transfer operators. *Nonlinear Processes in Geophysics*, 16(6):655–663.

Dellnitz, M., Froyland, G., and Junge, O. (2001). The algorithms behind gaio - set oriented numerical methods for dynamical systems. In *Ergodic theory, analysis, and efficient simulation of dynamical systems*, pages 145–174. Springer.

Deza, J. I. (2015). *Climate networks constructed by using information-theoretic measures and ordinal time-series analysis*. PhD Thesis. `http://www.tdx.cat/handle/10803/286281`.

Deza, J. I., Masoller, C., and Barreiro, M. (2013). Distinguishing the effects of internal and forced atmospheric variability in climate networks. *Nonlinear Processes in Geophysics*, (November).

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271.

Dobricic, S. (2005). New mean dynamic topography of the Mediterranean calculated from assimilation system diagnostic. *Geophys. Res. Lett.*, 32(11):L11606.

Dobricic, S. and Pinardi, N. (2008). An oceanographic three-dimensional variational data assimilation scheme. *Ocean Model.*, 22(3–4):89–105.

Dobricic, S., Pinardi, N., Adani, M., Tonani, M., Fratianni, C., Bonazzi, A., and Fernandez, V. (2007). Daily oceanographic analyses by mediterranean forecasting system at the basin scale. *Ocean Science*, 3(1):149–157.

Donges, J. F., Schultz, H. C., Marwan, N., Zou, Y., and Kurths, J. (2011a). Investigating the topology of interacting networks. *The European Physical Journal B*, 84(4):635–651.

Donges, J. F., Schultz, H. C. H., Marwan, N., Zou, Y., and Kurths, J. (2011b). Investigating the topology of interacting networks-theory and application to coupled climate subnetworks. *Eur. Phys. J. B*, 84:635–651.

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J. (2009a). The backbone of the climate network. *EPL (Europhysics Letters)*, 87(4):48007.

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J. (2009b). Complex networks in climate dynamics. *The European Physical Journal-Special Topics*, 174(1):157–179.

d'Ortenzio, F. and d'Alcalá, M. R. (2009). On the trophic regimes of the Mediterranean sea: a satellite analysis. *Biogeosc.*, 6:1–10.

d'Ovidio, F., Fernandez, V., Hernandez-García, E., and Lopez, C. (2004). Mixing structures in the Mediterranean Sea from Finite-Size Lyapunov Exponents. *Geophys. Res. Lett.*, 31:L17203.

Drévillon, M., Bourdalle-Badie, R., Derval, C., Drillet, Y., Lellouche, J.-M., Remy, E., Tranchant, B., Benkiran, M., Greiner, E., Guinehut, S., Verbrugge, N., Garric, G., Testut, C.-E., Laborie, M., Nouel, L., Bahurel, P., Bricaud, C., Crosnier, L., Dombrowsky, E., Durand, E., Ferry, N., Hernandez, F., Le-Galloudec, O., Messal, F., and Parent, L. (2008). The GODAE/Mercator ocean global ocean forecasting system: results, applications and prospects. *J. Operational Ocean.*, 1:51–57.

Dubois, M., Rossi, V., Ser-Giacomi, E., Arnaud-Haond, S., Lopez, C., and Hernandez-Garcia, E. (2015). Linking basin-scale connectivity, oceanography and population dynamics for the conservation and management of marine ecosystems. *Under review, Global Ecology and Biogeography*.

Fiedler, M. (1975). A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25(4):619–633.

Fortunato, S. (2010). Community detection in graphs. *Phys. Rep.*, 486:75–174.

Fortunato, S. and Barthélemy, M. (2007). Resolution limit in community detection. *P. Natl. Acad. Sci. USA*, 104:36–41.

Franchini, P., Sola, L., Crosetti, D., Milana, V., and Rossi, A. (2012). Low levels of population genetic structure in the gilthead sea bream, sparus aurata, along the coast of Italy. *ICES Journal of Marine Science: Journal du Conseil*, 69(1):41–50.

Froyland, G. (2005). Statistically optimal almost-invariant sets. *Physica D: Nonlinear Phenomena*, 200(3-4):205–219.

Froyland, G. and Dellnitz, M. (2003). Detecting and locating near-optimal almost-invariant sets and cycles. *SIAM Journal on Scientific Computing*, 24(6):1839–1863.

Froyland, G., Horenkamp, C., Rossi, V., Santitissadeekorn, N., and Gupta, A. S. (2012). Three-dimensional characterization and tracking of an Agulhas ring. *Ocean Modelling*, 52:69–75.

Froyland, G. and Padberg, K. (2009). Almost-invariant sets and invariant manifolds-connecting probabilistic and geometric descriptions of coherent structures in flows. *Physica D: Nonlinear Phenomena*, 238(16):1507–1523.

Froyland, G., Padberg, K., England, M. H., and Treguier, A. M. (2007). Detection of coherent oceanic structures via transfer operators. *Physical Review Letters*, 98(22):224503.

Froyland, G. and Padberg-Gehle, K. (2012). Finite-time entropy: A probabilistic approach for measuring nonlinear stretching. *Physica D: Nonlinear Phenomena*, 241(19):1612–1628.

Froyland, G., Santitissadeekorn, N., and Monahan, A. (2010). Transport in time-dependent dynamical systems: Finite-time coherent sets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4):–.

Froyland, G., Stuart, R. M., and van Sebille, E. (2014). How well-connected is the surface of the global ocean? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(3):–.

Gaggiotti, O. (1996). Population genetic models of source–sink metapopulations. *Theoretical Population Biology*, 50(2):178 – 208.

Game, E. T., Grantham, H. S., Hobday, A. J., Pressey, R. L., Lombard, A. T., Beckley, L. E., Gjerde, K., Bustamante, R., Possingham, H. P., and Richardson, A. J. (2009). Pelagic protected areas: the missing dimension in ocean conservation. *Trends Ecol. Evol.*, 24(7):360–369.

Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.

Guez, O., Gozolchiani, A., Berezin, Y., Brenner, S., and Havlin, S. (2012). Climate network structure evolves with North Atlantic oscillation phases. *EPL (Europhysics Letters)*, 98(3):38006.

Guidetti, P., Notarbartolo-Di-Sciara, G., and Agardy, T. (2013). Integrating pelagic and coastal MPAs into large-scale ecosystem-wide management. *Aquatic Conserv.: Mar. Freshw. Ecosyst.*, 23:179–182.

Guizien, K., Belharet, M., Marsaleix, P., and Guarini, J. (2012). Using larval dispersal simulations for marine protected area design: Application to the gulf of lions (northwest mediterranean). *Limnology and Oceanography*, 57(4):1099.

Haller, G. (2000). Finding finite-time invariant manifolds in two-dimensional velocity fields. *Chaos*, 10(1):99–108.

Haller, G. (2001). Distinguished material surfaces and coherent structure in three-dimensional fluid flows. *Physica D*, 149:248–277.

Haller, G. and Yuan, G. (2000). Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica D*, 147:352–370.

Hamdi, E., Goñi, R., Dìaz, D., and Planes, S. (2012). Detecting immigrants in a highly genetically homogeneous spiny lobster population (palinurus elephas) in the northwest Mediterranean sea. *Ecology and evolution*, 2(10):2387–2396.

Hamilton, S., Caselle, J., Malone, D., and Carr, M. (2010). Incorporating biogeography into evaluations of the Channel Islands marine reserve network. *P. Natl. Acad. Sci. USA*, 107:18272–18277.

Hartl, D. and Clark, A. (1998). *Principles of population genetics*. {Sinauer Associates}.

Hastings, A. and Botsford, L. (2006). Persistence of spatial populations depends on returning home. *Proceedings of the National Academy of Sciences*, 103(15):6067–6072.

Hernandez-Carrasco, I. (2013). *Horizontal transport and mixing and their connection with dynamical and biologic processes in the ocean*. PhD Thesis. `https://ifisc.uib-csic.es/publications/publication-detail.php?indice=2422`.

Hernández-Carrasco, I., López, C., Hernández-García, E., and Turiel, A. (2011). How reliable are finite-size Lyapunov exponents for the assesment of ocean dynamics? *Ocean Modelling*, 36(3-4):208–218.

Hernández-Carrasco, I., López, C., Hernández-García, E., and Turiel, A. (2012). Seasonal and regional characterization of horizontal stirring in the global ocean. *J. Geophys. Res.*, 117:C10007.

Hixon, M., Pacala, S., and Sandin, S. (2002). Population regulation: historical context and contemporary challenges of open vs. closed systems. *Ecology*, 83(6):1490–1508.

Hogan, J., Thiessen, R., Sale, P., and Heath, D. (2012). Local retention, dispersal and fluctuating connectivity among populations of a coral reef fish. *Oecologia*, 168(1):61–71.

Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3):97–125.

Jolly, M., Guyard, P., Ellien, C., Gentil, F., Viard, F., Thiébaut, E., and Jollivet, D. (2009). Population genetics and hydrodynamic modeling of larval dispersal dissociate contemporary patterns of connectivity from historical expansion into european shelf seas in the polychaete pectinaria koreni. *Limnology and Oceanography*, 54(6):2089.

Joseph, B. and Legras, B. (2002). Relation between Kinematic Boundaries, Stirring, and Barriers for the Antartic Polar Vortex. *J. Atm. Sci.*, 59:1198–1212.

Kaplan, D., Chassot, E., Gruss, A., and Fonteneau, A. (2010a). Pelagic MPAs: the devil is in the details. *Trends Ecol. Evol.*, 25:62–63.

Kaplan, D., Planes, S., Fauvelot, C., Brochier, T., Lett, C., Bodin, N., Le-Loch, F., Tremblay, Y., and Georges, J.-Y. (2010b). New tools for the spatial management of living marine resources. *Curr. Opin. Environ. Sust.*, 2:88–93.

Kim, H. and Anderson, R. (2012). Temporal node centrality in complex networks. *Physical Review E*, 85(2):026107.

Kool, J., Moilanen, A., and Treml, E. (2013). Population connectivity: recent advances and new perspectives. *Landscape Ecol.*, 28:165–185.

Kourafalou, V. and Barbopoulos, K. (2003). High resolution simulations on the North Aegean Sea seasonal circulation. *Ann. Geophys.*, 21:251–265.

Lacorata, G., Aurell, E., and Vulpiani, A. (2001). Drifter dispersion in the Adriatic sea: Lagrangian data and chaotic model. *Ann. Geophysicae*, 19:121–129.

Lagabrielle, E., Bissery, C., Crochelet, E., Meola, B., Webster, C., Claudet, J., Chassanite, A., Marinesque, S., Robert, P., Goutx, M., and Quod, C. (2012). The status of Marine Protected Areas in the Mediterranean Sea. Technical report, MedPAN Collection, CAR/ASP.

Lancichinetti, A. and Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Phys. Rev. E*, 80:056117.

Ledoux, J., Mokhtar-Jamaï, K., Roby, C., Féral, J., Garrabou, J., and Aurelle, D. (2010). Genetic survey of shallow populations of the mediterranean red coral [corallium rubrum (linnaeus, 1758)]: new insights into evolutionary processes shaping nuclear diversity and implications for conservation. *Molecular ecology*, 19(4):675–690.

Lenfant, P. and Planes, S. (1996). Genetic differentiation of white sea bream within the Lion's gulf and the Ligurian sea (Mediterranean sea). *Journal of Fish Biology*, 49(4):613–621.

Lentz, H. H., Selhorst, T., and Sokolov, I. M. (2013). Unfolding accessibility provides a macroscopic approach to temporal networks. *Physical Review Letters*, 110(11):118701.

Lester, S., Halpern, B., Grorud-Colvert, K., Lubchenco, J., Ruttenberg, B., Gaines, S., Airamé, S., and Warner, R. R. (2009). Biological effects within no-take marine reserves: a global synthesis. *Mar. Ecol. Prog. Ser.*, 384:33–46.

Lett, C., Nguyen-Huu, T., Cuif, M., Saenz-Agudelo, P., and Kaplan, D. (2015). Linking local retention, self-recruitment and persistence in marine metapopulations. *Ecology*, page in press.

Lindner, M. (2015). *A Markov chain approach to describing geophysical transport processes from a complex networks perspective*. Bachelor Thesis, Humboldt Univ. Berlin.

Longhurst, A. R. (2006). *Ecological Geography of the Sea*. Academic Press.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141.

Lowe, W. and Allendorf, F. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, 19(15):3038–3051.

Macpherson, E. and Raventos, N. (2006). Relationship between pelagic larval duration and geographic distribution of Mediterranean littoral fishes. *Mar. Ecol. Prog. Ser.*, 327:257–265.

Madec, G., Bissery, C., Crochelet, E., Meola, B., Webster, C., Claudet, J., Chassanite, A., Marinesque, S., Robert, P., Goutx, M., Quod, C., Robert, P., Goutx, M., and Quod, C. (2008). Nemo ocean engine. Technical report, Note du Pole de Modélisation, Institut Pierre-Simon Laplace (IPSL).

Mancho, A. M., Hernandez-Garcia, E., Small, D., Wiggins, S., and Fernandez, V. (2008). Lagrangian transport through an ocean front in the North-Western Mediterranean sea. *Journal of Physical Oceanography*, 38:1222–1237.

Mancho, A. M., Small, D., and Wiggins, S. (2004). Computation of hyperbolic trajectories and their stable and unstable manifolds for oceanographic flows represented trajectories. 11:17–33.

Mancho, A. M., Small, D., Wiggins, S., and Ide, K. (2003). Computation of stable and unstable manifolds of hyperbolic trajectories in two-dimensional, aperiodically time-dependent vector fields. *Physica D*, 182:188–222.

Matsueda, M. (2011). Predictability of Euro-Russian blocking in summer 2010. *Geophys. Res. Lett.*, 38:L06801.

Mheen, M., Dijkstra, H. A., Gozolchiani, A., Toom, M., Feng, Q., Kurths, J., and Hernandez-Garcia, E. (2013). Interaction network based early warning indicators for the Atlantic MOC collapse. *Geophysical Research Letters*, 40(11):2714–2719.

Millot, C. and Taupier-Letage, I. (2005). Circulation in the Mediterranean sea. In *The Mediterranean Sea*, pages 29–66. Springer.

Mokhtar-Jamaï, K., Pascual, M., Ledoux, J., Coma, R., Féral, J., Garrabou, J., and Aurelle, D. (2011). From global to local genetic structuring in the red gorgonian paramuricea clavata: the interplay between oceanographic conditions and limited larval dispersal. *Molecular Ecology*, 20(16):3291–3305.

Newman, M. (2009). *Networks: An introduction*. Oxford University Press.

Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113.

Nilsson-Jacobi, M., André, C., Doos, K., and Jonsson, P. R. (2012). Identification of subpopulations from connectivity matrices. *Ecography*, 35:1004–1016.

Oddo, P., Adani, M., Pinardi, N., Fratianni, C., Tonani, M., Pettenuzzo, D., et al. (2009). A nested Atlantic-Mediterranean sea general circulation model for operational forecasting. *Ocean Science*, 5(4):461.

Pala, C. (2013). Giant marine reserves pose vast challenges. *Science*, 339(6120):640–641.

Paladin, G. and Vulpiani, A. (1987). Anomalous scaling laws in multifractal objects. *Physics Reports*, 156(4):147 – 225.

Palumbi, S. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological applications*, 13(sp1):146–158.

Pannell, J. and Charlesworth, B. (1999). Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution*, pages 664–676.

Pelc, R. A., Warner, R. R., Gaines, S. D., and Paris, C. B. (2010). Detecting larval export from marine reserves. *P. Natl. Acad. Sci. USA*, 107(43):18266–18271.

Pikitch, E., Santora, C., Babcock, E., Bakun, A., Bonfil, R., Conover, D., Dayton, P., Doukakis, P., Fluharty, D., Heneman, B., Houde, E., Link, J., Livingston, P., Mangel, M., McAllister, M., Pope, J., and Sainsbury, K. (2004). Ecology: ecosystem-based fishery management. *Science*, 305:346–347.

Pinardi, N., Zavatarelli, M., Adani, M., Coppini, G., Fratianni, C., Oddo, P., Simoncelli, S., Tonani, M., Lyubartsev, V., Dobricic, S., and Bonaduce, A. (2013). Mediterranean sea large-scale low-frequency ocean variability and water mass formation rates from 1987 to 2007: A retrospective analysis. *Progress in Oceanography*, page in press.

Pineda, J., Hare, J., and Sponaugle, S. (2007). Larval transport and dispersal in the coastal ocean and consequences for population connectivity. *Oceanog.*, 20(3):22–39.

Planes, S., Jones, G., and Thorrold, S. (2009). Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences*, 106(14):5693–5697.

Poulos, S., Drakopoulos, P., and Collins, M. (1997). Seasonal variability in sea surface oceanographic conditions in the Aegean sea (eastern Mediterranean): an overview. *J. Mar. Syst.*, 13:225–244.

Pulliam, H. (1988). Sources, sinks, and population regulation. *American naturalist*, pages 652–661.

Rényi, A. (1970). *Probability Theory*. North-Holland, Amsterdam.

Ribeiro, B., Perra, N., and Baronchelli, A. (2013). Quantifying the effect of temporal resolution on time-varying networks. *Scientific Reports*, 3.

Roberts, C. (1997). Connectivity and management of caribbean coral reefs. *Science*, 278(5342):1454–1457.

Rossi, V., Ser-Giacomi, E., López, C., and Hernández-García, E. (2014). Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves. *Geophysical Research Letters*, 41(8):2883–2891.

Rosvall, M. and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123.

Rozenfeld, A., Arnaud-Haond, S., Hernández-García, E., Eguíluz, V. M., Serrão, E., and Duarte, C. (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences*, 105(48):18824–18829.

Saenz-Agudelo, P., Jones, G., Thorrold, S., and Planes, S. (2011). Connectivity dominates larval replenishment in a coastal reef fish metapopulation. *Proceedings of the Royal Society B: Biological Sciences*, page rspb20102780.

Saha, S. and coauthors (2010). The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, 91(8):1015–1057.

Santitissadeekorn, N., Froyland, G., and Monahan, A. (2010). Optimally coherent sets in geophysical flows: A transfer-operator approach to delimiting the stratospheric polar vortex. *Physical Review E*, 82(5):056311.

Schaub, M., Lambiotte, R., and Barahona, M. (2012). Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation. *Phys. Rev. E*, 86:026112.

Schneider, J., Fernández, V., and Hernández-García, E. (2005). Leaking method approach to surface transport in the Mediterranean Sea from a numerical ocean model. *J. Mar. Syst.*, 57:111–126.

Schneider, J. and Tél, T. (2003). Extracting flow structures from tracer data. *Ocean Dynamics*, 53(2):64–72.

Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintoré, J., Vidal-Vijande, E., Pascual, A., Guidetti, P., and Pascual, M. (2011). Matching genetics with oceanography: directional gene flow in a Mediterranean fish species. *Molecular ecology*, 20(24):5167–5181.

Ser-Giacomi, E., Rossi, V., López, C., and Hernández-García, E. (2015a). Flow networks: A characterization of geophysical fluid transport. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):036404.

Ser-Giacomi, E., Vasile, R., Hernández-García, E., and López, C. (2015b). Most probable paths in temporal weighted networks: An application to ocean transport. *Physical review E*, 92:036404.

Ser-Giacomi, E., Vasile, R., Recuerda, I., Hernández-García, E., and López, C. (2015c). Dominant transport pathways in an atmospheric blocking event. *Chaos: an interdisciplinary journal of nonlinear science*, 25:087413.

Shadden, S. C., Lekien, F., and Marsden, J. E. (2005). Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. *Physica D*, 212(34):271–304.

Shanks, A. L. (2009). Pelagic Larval Duration and dispersal distance revisited. *Biol. Bull.*, 216:373–385.

Siegel, D. A., Mitarai, S., Costello, C. J., Gaines, S. D., Kendall, B. E., Warner, R. R., and Winters, K. B. (2008). The stochastic nature of larval connectivity among nearshore marine populations. *P. Natl. Acad. Sci. USA*, 105(26):8974–8979.

Soria, G., Munguía-Vega, A., Marinone, S., Moreno-Báez, M., Martínez-Tovar, I., and Cudney-Bueno, R. (2012). Linking bio-oceanography and population genetics to assess larval connectivity. *Marine Ecology Progress Series*, 463:159–175.

Speetjens, M., Lauret, M., Nijmeijer, H., and Anderson, P. (2013). Footprints of lagrangian flow structures in eulerian concentration distributions in periodic mixing flows. *Physica D: Nonlinear Phenomena*, 250:20–33.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G. (2005). Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. *Atmos. Chem. Phys.*, 5:2461–2474.

Stohl, A., Sodemann, H., Eckhardt, S., Frank, A., Seibert, P., and Wotawa, G. (2011). The Lagrangian particle dispersion model FLEXPART version 8.2. *FLEXPART user guide*.

Tantet, A., Dijkstra, H. A., et al. (2014). An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature. *Earth System Dynamics*, 5(1):1–14.

Teacher, A., Andre, C., Jonsson, P., and Merila, J. (2013). Oceanographic connectivity and environmental correlates of genetic structuring in atlantic herring in the baltic sea. *Evolutionary Applications*, 6:549–567.

Tél, T. and Gruiz, M. (2006). *Chaotic dynamics: An introduction based on classical mechanics*. Cambridge Univ. Press, Cambridge.

Thomas, C. J., Lambrechts, J., Wolanski, E., Traag, V., Blondel, V. D., Deleersnijder, E., and Hanert, E. (2014). Numerical modelling and graph theory tools to study ecological connectivity in the Great Barrier Reef. *Ecol. Modell.*, 272:160–174.

Tonani, M., Pinardi, N., M.Adani, Bonazzi, A., Coppini, G., De-Dominicis, M., Dobricic, S., M.Drudi, Fabbroni, N., Fratianni, C., Grandi, A., Lyubartsev, S., Oddo, P., Pettenuzzo, D., Pistoia, J., and I.Pujol (2008). The Mediterranean ocean forecasting system, coastal to global operational oceanography: Achievements and challenges. In Dahlin, H., Bell, M. J., Fleming, N. C., and Pietersson, S., editors, *Proceedings of the Fifth International Conference on EuroGOOS*, number 28, Sweden. EuroGOOS Publication.

Treml, E., Roberts, J., Chao, Y., Halpin, P. N., Possingham, H., and Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integr. Comp. Biol.*, 52(4):525–537.

Tsonis, A. and Roebber, P. (2004). The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, 333:497–504.

Tsonis, A. and Swanson, K. (2008a). Topology and Predictability of El Niño and La Niña Networks. *Physical Review Letters*, 100(22):1–4.

Tsonis, A. A. and Swanson, K. L. (2008b). Topology and predictability of El Nino and La Nina networks. *Physical Review Letters*, 100(22):228502.

Tsonis, A. A., Swanson, K. L., and Roebber, P. J. (2006). What Do Networks Have to Do with Climate? *Bulletin of the American Meteorological Society*, 87(5):585–595.

Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A., and Costa, L. D. F. (2010). Community structure and dynamics in climate networks. *Climate Dynamics*, 37(5-6):933–940.

Vallis, G. K. (2006). *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press.

Vaz, A. C., Richards, K. J., Jia, Y., and Paris, C. B. (2013). Mesoscale flow variability and its impact on connectivity for the island of Hawaii. *Geophys. Res. Lett.*, 40:332337.

White, C., Selkoe, K., Watson, J., Siegel, D. A., Zacherl, D., and Toonen, R. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences*.

Yamasaki, K., Gozolchiani, A., and Havlin, S. (2008). Climate networks around the globe are significantly affected by El Niño. *Physical Review Letters*, 100(22):228501.

Yamasaki, K., Gozolchiani, A., Havlin, S., and Gan, R. (2009). Climate Networks Based on Phase Synchronization Analysis Track El-Niño. *Progress of Theoretical Physics Supplement*, (179):178–188.

CURRICULUM VITAE

ENRICO SER-GIACOMI

Personal Details:

Date of birth: 10th November, 1986.
Place of birth: Perugia, Italy.
Mobile Phone: 0034 628932818
Nationality: Italian.
e-mail: enrico.sergiacomi@gmail.com
Skype: enrico.sergiacomi
Google Scholar Profile: Enrico Ser-Giacomi (visit)

Research Interests and Expertise:

DYNAMICAL SYSTEMS

- Nonlinearity, bifurcations and chaos
- Advection in chaotic flows

NETWORK THEORY

- Statistical properties, local and global analysis of complex networks
- Temporal, weighted, and space-embedded networks
- Conservation rules and symmetries in networks built from physical connections

GEOPHYSICS

- Atmospheric and oceanic fluid dynamics
- Climate dynamics
- Lagrangian transport phenomena

Current Position:

Ph.D. Student (Marie-Curie ITN program).
Center: University of Balearic Islands.
Institute: IFISC (CSIC-UIB) - Institute for Cross-Disciplinary Physics and Complex Systems.
Dedication: Full time.

Education:

10/2008–05/2011	Master degree in Theoretical Physics - 110/110 with honors. (read thesis) University of Perugia (Italy).
10/2005–10/2008	Bachelor degree in Physics - 110/110 with honors. University of Perugia (Italy).

Research Stays Abroad (more than 4 weeks):

09/2013–10/2013	IMAU - Institute for Marine and Atmospheric research Utrecht. Utrecht, The Netherlands.
06/2014–07/2014	VORtech - `http://www.vortech.nl`. Delft, The Netherlands.

Short Visits:

02/2014	IMAU - Institute for Marine and Atmospheric research Utrecht. Utrecht, The Netherlands.
06/2015	PIK - Potsdam Institute for Climate Impact Research. Potsdam, Germany.

Research Projects:

2012–2015	Learning about Interacting Networks in Climate: LINC. *Type of project:* European Commission Marie-Curie ITN (FP7- 320 PEOPLE-2011-ITN). *Partners:* 6 universities and 3 companies.

People: 12 members, 16 recruited researchers (from 12 countries).
Total ammount: 3.715.523,90 euros .
P.I.: Prof. Cristina Masoller (Universitat Politecnica de Catalunya).

Publications in Peer-Reviewed Journals:

1. D Astolfi, G Grignani, E Ser-Giacomi, AV Zayakin
   Strings in $AdS_4 \times CP^3$*: finite size spectrum vs. Bethe Ansatz*
   Journal of High Energy Physics **4**, 1-32 (2012)

2. V Rossi, E Ser-Giacomi, C Lopez, E Hernandez-Garcia
   Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves
   Geophysical Research Letters **41**, 2883-2891(2014)

   - Covered by national press, El Mundo. (read article)
     El mundo - Baleares, 15 April 2014

3. E Ser-Giacomi, V Rossi, C Lopez, E Hernandez-Garcia
   Flow networks: A characterization of geophysical fluid transport
   Chaos **25**, 036404 (2015).

4. T Squartini, E Ser-Giacomi, D Garlaschelli, G Judge
   Information Recovery In Behavioral Networks
   PLoS ONE 10(5): e0125077 (2015).

5. E Ser-Giacomi, R Vasile, C Lopez, E Hernandez-Garcia
   Most probable paths in temporal weighted networks: An application to ocean transport
   Physical Review E **92**, 012818 (2015)

6. E Ser-Giacomi, R Vasile, I Recuerda, E Hernandez-Garcia, C Lopez
   Dominant transport pathways in an atmospheric blocking event
   Chaos **25**, 087413 (2015).

Submitted manuscripts:

- M Dubois, V Rossi, E Ser-Giacomi, S Arnaud-Haond, C Lopez, E Hernandez-Garcia
  Linking basin-scale connectivity, oceanography and population dynamics for the conservation and management of marine ecosystems.
  Under review, Global Ecology and Biogeography (2015).

Participation in Conferences, Seminars and Workshops:

ORAL PRESENTATIONS:

- LINC Workshop. Soesterberg (The Netherlands). 25 April 2013. *Networks of oceanic transport*
- Complex Networks and Climate - research and scientific software tools. VORtech, Delft (The Netherlands). 18 October 2013. *Flow transport networks*
- LINC Mid-Term Review. PIK, Potsdam (Germany). 20 November 2013. *Lagrangian transport networks*
- LINC Workshop. Facultad de Ciencias, Montevideo (Uruguay). 24-25 March 2014. *Flow transport networks: from Mediterranean to North-Atlantic*
- IFISC Seminary. IFISC, Palma de Mallorca (Spain). 4 November 2014. *Flow networks: characterizing transport and connectivity in the ocean*
- International school - Mathematical methods: applications to earth science. Lomonosov Moscow State University, Moscow (Russia). 15-19 November 2014. *Flow transport network: an application to the Mediterranean Sea*
- Complex Systems and Climate Variability. Mercure Hotel, Vienna (Austria). 10-12 April 2015. *Lagrangian Flow Networks*
- HYDROGENCONNECT Workshop. IFISC, Palma de Mallorca (Spain). 29-30 June 2015. *A network characterization of oceanic connectivity and its potential application to ecology*
- CONFLOW 2015: Complex network perspectives on flow systems. PIK, Potsdam (Germany). 21-22 September 2015. *Most probable paths in time-dependent flow networks*

POSTERS:

- XXXIII Dynamics Days Europe. Madrid (Spain). 3-7 June 2013. *Networks of Lagrangian transport in the Mediterranean Sea*
- WCRP Conference for Latin America and the Caribbean. Montevideo (Uruguay). 17-21 March 2014. *Geophysical-flow transport networks*
- European Geosciences Union Assembly 2015. Vienna, (Austria). 13-17 April 2015. *Flow networks: a new to characterize transport and connectivity in geophysical flows*

ATTENDANCE:

- IV Summer School on Statistical Physics of Complex and Small Systems. IFISC, Palma de Mallorca (Spain). 8-19 September 2014.

- Complex Networks and Climate - research and scientific software tools. VORtech, Delft (The Netherlands). 18 October 2013.
- European Geosciences Union Assembly 2013. Vienna, (Austria). 7-12 April 2013.
- FisEs 2012. Palma de Mallorca (Spain). 18-20 October 2012
- Nonlinear Processes in Oceanic and Atmospheric Flows. Madrid (Spain). 3-6 July 2012.
- LACES 2010. Galileo Galilei Institute for Theoretical Physics, Arcetri (Italy). 22 November - 10 December 2010

Organization of Conferences and Workshops:

- CONFLOW 2015: Complex network perspectives on flow systems. PIK, Potsdam (Germany). 21-22 September 2015.

Mentorship Experience:

07/2014–08/2014	Co-supervisor of an Intro-IFISC 2014 grant. Jose Antonio Sanchez Pelegrin (Undergraduate student). Grant: SURF@IFISC 2014.

Teaching Experience:

02/2015–06/2015	Teaching assistant with Prof. Claudio Mirasso: Physics for Biology (University of Balearic Islands).

Computer skills and competences:

- Programming/markup languages: C/C++, Python, Matlab, LaTeX.
- Operating systems: Linux, Mac, Windows.
- High performance computing.

Language Skills:

Italian: Mother tongue.

English: Proficient Communication.

Spanish: Proficient Communication.

Society Membership:

2015	European Geosciences Union.

Updated on November 9, 2015.